EMPIRICAL BAYESIAN TEST OF THE SMOOTHNESS

By Eduard Belitser and Farida Enikeeva

Utrecht University and EURANDOM

7th April 2005

In the context of adaptive nonparametric curve estimation problem, a common assumption is that a function (signal) to estimate belongs to a nested family of functional classes, parameterized by a quantity which often has a meaning of smoothness amount. It has already been realized by many that the problem of estimating the smoothness is not sensible. What then can be inferred about the smoothness? The paper attempts to answer this question. We consider the implications of our results to hypothesis testing. We also relate them to the problem of adaptive estimation. The test statistic is based on the marginalized maximum likelihood estimator of the smoothness for an appropriate prior distribution on the unknown signal.

1 Introduction

Suppose we observe independent Gaussian data $X = (X_i)_{i\in\mathbb{N}}$, where $X_i \sim \mathcal{N}(\theta_i, n^{-1})$, $\theta = (\theta_i)_{i\in\mathbb{N}} \in \ell_2$ is an unknown parameter. This model is the sequence version of the Gaussian white noise model $dY(t) = f(t)dt + n^{-1/2}dW(t)$, $t \in [0,1]$, where $f \in \mathcal{L}_2[0,1] = \mathcal{L}_2$ is an unknown signal and W is the standard Brownian motion. The infinite dimensional parameter $\theta \in \ell_2$ can be regarded as the sequence of the Fourier coefficients of $f \in \mathcal{L}_2$ with respect to some orthonormal basis in $\mathcal{L}_2$. Sometimes we will call θ a signal.

The white noise model has received much attention in the last few decades and comprehensive treatments of it can be found in Ibragimov and Khasminski (1981) and Johnstone (1999).

Besides of being of interest on its own (the problem of recovering a signal transmitted over a communication channel with Gaussian white noise of intensity $n^{-1/2}$), the white noise model turns out to be a mathematical idealization of some other nonparametric models. For instance, the white noise model arises as the limiting experiment as $n \to \infty$, for the model of n i.i.d. observations with unknown density (see Nussbaum (1996), Grama and Nussbaum (1998)) and for the regression model (see Brown and Low (1996)). On the other hand, this model captures the statistical essence of the original model and preserves its main features in a pure form; cf. Johnstone (1999). Most of the estimation problems are studied in asymptotic setup ($n \to \infty$) from the viewpoint of increasing information. In

AMS 2000 subject classification. Primary 62G10; secondary 62C10, 62G20, 62C12.
Key words and phrases. Empirical Bayes, hypothesis testing, smoothness parameter, white noise model.
Abbreviated Title. Empirical Bayesian test.

fact, one deals with a sequence of models parameterized by n. Though non-asymptotic estimation problems are also very important, they are often not tractable mathematically. Our approach in this paper is also primarily asymptotic, however the intention is to derive non-asymptotic results as well, to be able to precisely evaluate the influence of different quantities and constants on the quality of the inference.

An enormous number of problems have already been studied involving statistical inference about a signal as in the white noise model. For example, signal estimation under different norms, estimation of a functional of the signal, testing hypothesis about the signal, construction of confidence sets.

A typical approach to the problems mentioned above is to assume that the unknown vector θ belongs to some compact set $\Theta_\beta \subset \ell_2$ indexed by $\beta \in \mathcal{B}$ which has a meaning of smoothness (here we consider only one-dimensional $\beta \in \mathcal{B} \subseteq \mathbb{R}$). If the parameter β is known, then we can use this knowledge in making inference about θ. If this knowledge is not available, an adaptation problem arises. For example, in the problem of adaptive estimation of θ, we have a list of models $\{\Theta_\beta\}$, $\beta \in \mathcal{B}$. It is then desirable to construct an estimator that depends only on the data X and asymptotically efficient for any subset Θ_β from the list.

Another way to look at the adaptation problem is based on the oracle inequalities approach (see Cavalier, Golubev, Picard and Tsybakov (2002), Tsybakov (2004)). Suppose that θ belongs to one of the compact subspaces Θ_β of the parameter space Θ and β is unknown. Let us choose a family of estimators $\mathcal{P}$, for instance, it can be the family of all possible projection estimators. Then for any fixed θ we can define an optimal (oracle) "estimator" from the family $\mathcal{P}$ which minimizes the risk over all the estimators from the family. Of course, the oracle is not really an estimator since it depends on θ that is unknown. The idea of oracle inequalities approach is to find an estimator $\widehat{\theta}$ depending only on the observations such that its risk can be well approximated from above by the risk of the oracle for any fixed $\theta \in \Theta$. This approximation is called oracle inequality and it upper bounds the risk of $\widehat{\theta}$. Moreover, this inequality provides us with a correct upper bound to the minimax risk over the set Θ_β. Indeed, since the risk of the oracle is minimal among all the estimators from $\mathcal{P}$, it will be less than the risk of the minimax projection estimator over the set Θ_β. Then the risk of the estimator $\widehat{\theta}$ will be also less than the risk of the minimax projection estimator over Θ_β. Thus our estimator $\widehat{\theta}$ is adaptive and gives correct minimax rate of convergence for any $\theta \in \Theta$. Although this method does provide us with an optimal adaptation procedure, we do not learn anything about the subspace Θ_β our parameter belongs to. In this paper we try to find a procedure that can give us information about the parameter class. Our goal is to find out what we can say about the smoothness of our parameter θ.

However, it is not quite clear how to characterize the amount of smoothness that a particular signal has. Suppose we are given that $\theta \in \Theta_\beta$ for some known $\beta \in \mathcal{B}$. Using this information, we can construct an estimator which is minimax over Θ_β. However, it can happen that the smoothness of θ is greater than β. For example, if the family of sets Θ_β is nested: $\Theta_{\beta_1} \subset \Theta_{\beta_2}$, $\beta_1 > \beta_2$, and $\mathcal{B}$ is a continuous set, then for any $\theta \in \Theta_{\beta_2}$ we can find $\beta_1 > \beta_2$ such that $\theta \in \Theta_{\beta_1}$. In this case the performance of our estimator under the assumption $\theta \in \Theta_\beta$ will not be the best possible. Thus even if we know that $\theta \in \Theta_\beta$, it is still more advantageous to use an adaptive estimator, since the above smoothness condition on θ does not exclude the case $\theta \in \Theta_{\beta_1}$ with $\beta_1 > \beta$. This actually means that for each particular $\theta \in \Theta_\beta$, it might be possible to find an adaptive procedure with better performance than the minimax one, which leads to the super-efficiency phenomenon. A good adaptation method should use this fact and provide better quality of the statistical inference corresponding to the smoothness of this particular signal.

Interestingly, this question is not relevant in the minimax setting. Indeed, in this case the measure of estimation quality involves taking the supremum over $\theta \in \Theta_\beta$ and therefore, according to this criteria, adaptive procedures can at best attain the minimax risk. An intuitive explanation is that, under the minimax criteria, all the estimation procedures are oriented to the worst signal (sequence of signals) from the class Θ_β with the smoothness at most β. Thus the generic smoothness is always the one which appears in the definition of the minimax risk. At first sight, adaptive procedures in the minimax sense may not seem very attractive, but, fortunately, in many cases they lead to superefficient estimators.

Many estimators considered in the literature are shown to be minimax or nearly minimax over some functional classes or some functional scales; see, for example, Donoho, Johnstone, Kerkyacharian, and Picard (1995) and further references therein. However, in practice only one signal is to be estimated. Practitioners would prefer an estimator with better performance for the underlying unknown signal rather than an estimator which is minimax over a functional class (or nearly minimax over a functional scale), unless one wants to estimate several signals by the same estimator.

The paper is organized as follows. Section 2 describes the empirical Bayes approach. The main results are given in Section 3. We prove some auxiliary lemmas in Section 4. In Section 5 we relate our results to the problem of testing the smoothness of the signal. In Section 6 we discuss some implications of the results to the problem of adaptive estimation of θ. Some technical lemmas are gathered in Appendix.

2 Empirical Bayes approach

Let $\{\Theta_\beta\}_{\beta\in\mathcal{B}}$, $\mathcal{B} = (1/2, +\infty)$ be a family of Sobolev type subspaces

$$\theta \in \Theta_\beta = \left\{\theta : \sum_{i=1}^{\infty} i^{2\beta}\theta_i^2 < \infty\right\}.$$

We suppose that θ belongs to a certain $\Theta_\beta \subset \ell_2$ for some unknown $\beta \in \mathcal{B}$.

For a particular $\theta \in \ell_2$ define a function

$$A_\theta(\beta) = \sum_{i=1}^{\infty} i^{2\beta}\theta_i^2\,, \qquad \beta \in \mathcal{B}\,. \tag{1}$$

It is a monotone function of β. Note that $\theta \in \Theta_\beta$ if and only if $A_\theta(\beta) < \infty$. Throughout the paper we assume that there exists $\bar\beta \in \mathcal{B}$ such that $\bar\beta = \bar\beta(\theta) = \sup\{\beta \in \mathcal{B} : \ A_\theta(\beta) < \infty\}$. We will call $\bar\beta \equiv \bar\beta(\theta)$ the smoothness of θ. Two possibilities may occur: either $A_\theta(\beta) \to \infty$

as $\beta \uparrow \bar{\beta}$ or $A_\theta(\bar{\beta}) < \infty$ and $A_\theta(\beta) = \infty$ for all $\beta > \bar{\beta}$. It is the behavior of this function $A_\theta(\beta)$ that effectively measures the smoothness of the underlying signal θ. Unless otherwise specified, we assume from now on that $A_\theta(\bar{\beta}) = \infty$.

The main goal of this paper is to make an inference about the smoothness of the signal on the basis of the observed data X. The inference will be based on a statistic $\hat{\beta}(X)$ (with the intuitive meaning of being an estimator of smoothness), which we construct using the empirical Bayes approach. In the next section we will make this problem mathematically formal by evaluating so-called probabilities of undersmoothing and oversmoothing for this statistic. In the rest of this section we describe the construction of $\hat{\beta}(X)$. The idea of the approach is to put a "right" prior $\pi(\beta)$ on the parameter θ, find the marginal distribution of X which will depend on β, and then use the marginal maximum likelihood estimator of β as the test statistic.

We need to clarify the meaning of being "right" for a family of priors $\pi(\beta)$, $\beta \in \mathcal{B}$. As it is well illustrated in a series of papers by Diaconis and Freedman, an arbitrary choice of the prior may lead to Bayesian procedures that easily fail in infinite dimensional problems. In order to be "right", the prior should reflect adequately the smoothness assumption on the unknown signal. There are many ways to describe this. Here we propose the following guiding principle, which adapts to the inference problem on θ. For example, the inference problems can be estimation of θ, estimation of a functional of θ, testing hypotheses, constructing confidence set. Usually these problems come with their own performance criteria, like the rate of convergence for the estimation problem. According to our principle, we require a prior that guarantees high performance relative to the given criteria under the given Bayesian and simultaneously under the corresponding frequentist formulation. For instance, in the case of an estimation problem, the Bayesian estimator should be a minimax estimator, at least with respect to the convergence rate. This principle should not be taken as a precise prescription, but rather as a starting point in the choice of "correct" priors in infinite dimensional statistical problems. In each particular statistical problem, one has to investigate the performance of the resulting Bayesian procedure. The choice of the prior surely depends on the underlying inference problem on θ.

In this paper, we consider the following version of the above principle: we take the underlying inference problem on θ to be the problem of estimating θ in ℓ_2-norm. Thus, we should choose a prior leading to a Bayes estimator that is at least rate optimal in the minimax sense over the corresponding class with smoothness β. The minimax ℓ_2-rate over the Sobolev ellipsoid of smoothness β is $n^{2\beta/(2\beta+1)}$ (see Pinsker (1980)) and the Bayes risk of our estimator should attain the same convergence rate. We put the following prior $\pi = \pi(\beta)$ on θ: the θ_i's are independent and for $\delta \in [0,1]$

$$\theta_i \sim \mathcal{N}(0, \tau_i^2(\beta)), \qquad \tau_i^2(\beta) = \tau_i^2(\beta, \delta, n) = n^{\frac{\delta-1}{2\beta+1}} i^{-(2\beta+\delta)}, \quad i \in \mathbb{N}. \tag{2}$$

Recall the following simple fact: if $Z|Y \sim \mathcal{N}(Y, \tau^2)$ and $Y \sim \mathcal{N}(\mu, \sigma^2)$, then

$$Y|Z \sim \mathcal{N}\Big(\frac{Z\sigma^2 + \mu\tau^2}{\tau^2 + \sigma^2}, \frac{\tau^2\sigma^2}{\tau^2 + \sigma^2}\Big).$$

Let E_π denote the expectation with respect to the prior π. The Bayesian estimator of

θ based on the above prior is the vector $\hat\theta = (\hat\theta_i)_{i\in\mathbb{N}}$ with components

$$\hat\theta_i = \hat\theta_i(\beta) = \mathrm{E}(\theta_i|X_i) = \frac{\tau_i^2(\beta)X_i}{\tau_i^2(\beta)+n^{-1}}\,, \quad i\in\mathbb{N}\,. \tag{3}$$

The choice of the prior and the variance (2) is made according to our principle as the following lemma shows.

We will need the following notation. For $0<p<\infty$, $0<q<\infty$, $0\le r<\infty$ such that $pq>r+1$ denote

$$B(p,q,r) = \int_0^\infty \frac{u^r}{(1+u^p)^q}\,du = p^{-1}\mathrm{Beta}\Big(q-\frac{r+1}{p},\frac{r+1}{p}\Big)\,, \tag{4}$$

where, for $\alpha,\beta>0$, $\mathrm{Beta}(\alpha,\beta)=\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du$ is the Beta function.

Lemma 1. *Let $\theta\in\Theta_\beta$ for some $\beta\in\mathcal{B}$ and $\hat\theta$ be defined by (3). Then, as $n\to\infty$,*

$$\mathrm{E}_\pi\|\theta-\hat\theta\|^2 = n^{-2\beta/(2\beta+1)}B(2\beta+\delta,1,0)(1+o(1))\,,$$

$$\mathrm{E}_\theta\|\theta-\hat\theta\|^2 \le n^{-2\beta/(2\beta+1)}\big(A_\theta(\beta)C(\beta,\delta)+B(2\beta+\delta,2,0)\big)(1+o(1))\,,$$

where $C(\beta,\delta)=\dfrac{(1+\delta\beta^{-1})^{2(\beta+\delta)/(2\beta+\delta)}}{(2+\delta\beta^{-1})^2}$ and function B is defined by (4).

Proof. By (2) and Lemma 8, we have, as $n\to\infty$,

$$\mathrm{E}_\pi\|\theta-\hat\theta\|^2 = \sum_{i=1}^\infty \frac{\tau_i^2(\beta)n^{-1}}{\tau_i^2(\beta)+n^{-1}} = \sum_{i=1}^\infty \frac{n^{-(1-\delta)(2\beta+1)}}{n^{(2\beta+\delta)(2\beta+1)}+i^{2\beta+\delta}} = n^{-2\beta/(2\beta+1)}B(2\beta+\delta,1,0)(1+o(1))\,.$$

The frequentist risk consists of two terms

$$\mathrm{E}_\theta\|\theta-\hat\theta\|^2 = \mathrm{E}_\theta\sum_{i=1}^\infty\Big(\frac{\tau_i^2(\beta)X_i}{\tau_i^2(\beta)+n^{-1}}-\theta_i\Big)^2 = \sum_{i=1}^\infty \frac{n^{-2}\theta_i^2}{(\tau_i^2(\beta)+n^{-1})^2} + \sum_{i=1}^\infty \frac{n^{-1}\tau_i^4(\beta)}{(\tau_i^2(\beta)+n^{-1})^2}\,.$$

Using again (2) and Lemma 8, we bound these terms as follows: as $n\to\infty$,

$$\begin{aligned}\sum_{i=1}^\infty \frac{n^{-2}\theta_i^2}{(\tau_i^2(\beta)+n^{-1})^2} &= \sum_{i=1}^\infty \frac{i^{2(2\beta+\delta)}\theta_i^2}{(n^{(2\beta+\delta)/(2\beta+1)}+i^{2\beta+\delta})^2} \le A_\theta(\beta)\max_{i\in\mathbb{N}}\frac{i^{2\beta+2\delta}}{\big(n^{(2\beta+\delta)/(2\beta+1)}+i^{2\beta+\delta}\big)^2}\\ &\le A_\theta(\beta)C(\beta,\delta)n^{-2\beta/(2\beta+1)}(1+o(1))\,;\end{aligned}$$

$$\sum_{i=1}^\infty \frac{n^{-1}\tau_i^4(\beta)}{(\tau_i^2(\beta)+n^{-1})^2} = \sum_{i=1}^\infty \frac{n^{(2\beta+2\delta-1)/(2\beta+1)}}{\big(n^{(2\beta+\delta)/(2\beta+1)}+i^{2\beta+\delta}\big)^2} = n^{-2\beta/(2\beta+1)}B(2\beta+\delta,2,0)(1+o(1))\,.$$

The lemma is proved. □

Below we present another lemma which justifies in a way the choice of variance of the prior distribution. Roughly speaking, this lemma says that if θ belongs to the set Θ_β, then the estimator $\widehat\theta$ belongs to the same set with probability tending to 1.

Lemma 2. *Let $\theta \in \Theta_\beta$ for $\beta > 1/2$. Then*

$$\lim_{T\to\infty} \sup_{n\geq 1} \mathrm{P}_\theta\Big\{\sum_{i=1}^{\infty} \widehat{\theta}_i^2 i^{2\beta} > T\Big\} = 0.$$

Proof. By Chebyshev's inequality,

$$\mathrm{P}_\theta\Big\{\sum_{i=1}^{\infty} \widehat{\theta}_i^2 i^{2\beta} > T\Big\} \leq T^{-1} \sum_{i=1}^{\infty} \mathrm{E}_\theta(\widehat{\theta}_i^2) i^{2\beta}.$$

Note that

$$\mathrm{E}_\theta(\widehat{\theta}_i^2) = \left(\frac{\tau_i^2(\beta)}{\tau_i^2(\beta)+n^{-1}}\right)^2 \mathrm{E}_\theta X_i^2 = n^{2(2\beta+\delta)/(2\beta+1)} \frac{\theta_i^2 + n^{-1}}{(n^{(2\beta+\delta)(2\beta+1)} + i^{2\beta+\delta})^2}.$$

Therefore,

$$\mathrm{P}_\theta\Big\{\sum_{i=1}^{\infty} i^{2\beta}\widehat{\theta}_i^2 > T\Big\} \leq T^{-1} \sum_{i=1}^{\infty} i^{2\beta} n^{2(2\beta+\delta)/(2\beta+1)} \frac{\theta_i^2 + n^{-1}}{(n^{(2\beta+\delta)/(2\beta+1)} + i^{2\beta+\delta})^2}$$
$$= T^{-1} n^{2(2\beta+\delta)(2\beta+1)} \left(\sum_{i=1}^{\infty} \frac{i^{2\beta}\theta_i^2}{(n^{(2\beta+\delta)/(2\beta+1)} + i^{2\beta+\delta})^2} + \sum_{i=1}^{\infty} \frac{i^{2\beta} n^{-1}}{(n^{(2\beta+\delta)/(2\beta+1)} + i^{2\beta+\delta})^2}\right).$$

It is sufficient to show that the sums in the above display are finite and bounded in n. Obviously, the first sum is bounded from above by $A_\theta(\beta)$. Applying Lemma 8 we can estimate the second sum. Thus we obtain

$$\mathrm{P}_\theta\Big\{\sum_{i=1}^{\infty} \widehat{\theta}_i^2 i^{2\beta} > T\Big\} \leq T^{-1}\left(A_\theta(\beta) + B(2\beta+\delta, 2, 2\beta)\right).$$

and prove the lemma. □

Remark 1. From now on we assume that $\delta = 1$, unless otherwise is specified.

Recall that we have the following marginal distribution of X: the X_i's are independent and $X_i \sim N\big(0, \tau_i^2(\beta) + n^{-1}\big)$, $i \in \mathbb{N}$. Let $L_n(\beta) = L_n(\beta, X)$ be the marginal likelihood of the data $X = (X_i)_{i\in\mathbb{N}}$:

$$L_n(\beta) = \prod_{i=1}^{\infty} \frac{1}{\sqrt{2\pi\big(\tau_i^2(\beta)+n^{-1}\big)}} \exp\Big\{-\frac{X_i^2}{2\big(\tau_i^2(\beta)+n^{-1}\big)}\Big\}.$$

Maximizing the function $L_n(\beta)$ is equivalent to minimizing $-\log L_n(\beta)$. To avoid complications in defining the minimum of $-\log L_n(\beta)$ under the event $\{-\log L_n(\beta) = \infty\}$, it is convenient to introduce $Z_n(\beta) = Z_n(\beta, \beta_0) = -2\log\frac{L_n(\beta)}{L_n(\beta_0)}$ for some reference value $\beta_0 \in \mathcal{B}$. For any set $S \subset \mathcal{B}$, define a marginal likelihood estimator of β restricted to the set S:

$$\widehat{\beta}(S) \equiv \widehat{\beta}(S, X, n) = \arg\min_{\beta\in S} Z_n(\beta). \tag{5}$$

This means that $Z_n(\hat\beta(S)) \le Z_n(\beta)$ for all $\beta \in S$ or, equivalently,

$$\sum_{i=1}^{\infty} \frac{\big(\tau_i^2(\beta_0) - \tau_i^2(\hat\beta(S))\big)X_i^2}{(\tau_i^2(\beta_0)+n^{-1})(\tau_i^2(\hat\beta(S))+n^{-1})} + \sum_{i=1}^{\infty} \log \frac{\tau_i^2(\hat\beta(S))+n^{-1}}{\tau_i^2(\beta_0)+n^{-1}}$$
$$\le \sum_{i=1}^{\infty} \frac{\big(\tau_i^2(\beta_0) - \tau_i^2(\beta)\big)X_i^2}{(\tau_i^2(\beta_0)+n^{-1})(\tau_i^2(\beta)+n^{-1})} + \sum_{i=1}^{\infty} \log \frac{\tau_i^2(\beta)+n^{-1}}{\tau_i^2(\beta_0)+n^{-1}}$$

for all $\beta \in S$. It follows also that $Z_n(\hat\beta(S), \beta) \le 0$ for all $\beta \in S$.

Denote for brevity

$$a_i = a_i(\beta,\beta_0) = \frac{1}{\tau_i^2(\beta)+n^{-1}} - \frac{1}{\tau_i^2(\beta_0)+n^{-1}} = \frac{\tau_i^2(\beta_0) - \tau_i^2(\beta)}{(\tau_i^2(\beta_0)+n^{-1})(\tau_i^2(\beta)+n^{-1})}, \tag{6}$$

$$b_i = b_i(\beta,\beta_0) = \frac{\tau_i^2(\beta)+n^{-1}}{\tau_i^2(\beta_0)+n^{-1}}. \tag{7}$$

Then $Z_n(\beta,\beta_0) = \sum_{i=1}^{\infty} a_i(\beta,\beta_0)X_i^2 + \sum_{i=1}^{\infty} \log b_i(\beta,\beta_0)$, and for all $\beta \in S$

$$\sum_{i=1}^{\infty} a_i(\hat\beta(S),\beta)X_i^2 \le \sum_{i=1}^{\infty} \log \big[b_i(\hat\beta(S),\beta)\big]^{-1}. \tag{8}$$

Remark 2. It is not so difficult to check that the above $\hat\beta$ can be related to a penalized least square estimator with the penalty $\mathrm{pen}(\theta,\beta) = \sum_{i=1}^{\infty} \big[\theta_i^2 \tau_i^{-2}(\beta) + \log\big(\tau_i^{-2}(\beta) + n^{-1}\big)\big]$. Indeed,

$$\hat\beta(S) = \arg\min_{\beta\in S} Z_n(\beta) = \arg\min_{\beta\in S}\min_{\theta\in\Theta_\beta} \Big\{ n\sum_{i=1}^{\infty}(X_i-\theta_i)^2 + \mathrm{pen}(\theta,\beta)\Big\}.$$

As to the set S, we assume it to be finite, dependent on n in such a way that $S = S_n$ forms an ε_n-net in $(1/2, \sup\{S_n\}]$, with $\varepsilon_n = o\big(1/(\log n)\big)$ and $\sup\{S_n\} \to \infty$ as $n \to \infty$. The requirement $\varepsilon_n = o\big(1/(\log n)\big)$ stems from the fact that $n^{2\beta_1/(2\beta_1+1)} = O\big(n^{2\beta_2/(2\beta_2+1)}\big)$ if $|\beta_1 - \beta_2| = O\big(1/(\log n)\big)$. For a set B, denote by $|B|$ the number of elements in the set B, possibly taking the value ∞. We also suppose that $|S_n| \le C_1 \exp\{C_2 n^\gamma\}$ for sufficiently large n for certain positive constants C_1, C_2, γ. Still there are many possible choices: for example, $\varepsilon_n = n^{-1}$ and $S_n = \{1/2 + k\epsilon_n,\ k = 0, 1, \ldots, n^2\}$ meet all the requirements.

Remark 3. We have chosen to maximize the process $Z_n(\beta)$ over some finite set $S = S_n$ to avoid unnecessary technical complications. Indeed, we could also take S to be the whole set $\mathcal{B}$ and then study the behavior of a (near) minimum point of $Z_n(\beta)$. The usual technique in such cases inspired by the empirical processes theory is to consider the minimum over some finite grid in $\mathcal{B}$ and to make sure at the same time that the increments of the process $Z_n(\beta)$ are uniformly small over small intervals (provided the process is smooth enough). We do not pursue this approach simply because it boils down to the same considerations as in the case when we restrict the minimization to the finite set S_n from the very beginning.

3 Main results

Let $\theta \in \Theta_{\beta_0}$. For any $\beta < \beta_0 \in S_n$ it is reasonable to call $\mathrm{P}_\theta\{\hat{\beta}(S_n) \le \beta\}$ the probability of undersmoothing. Given $\theta \in \Theta_{\beta_0}$, we would like to prove that $\mathrm{P}_\theta\{\hat{\beta}(S_n) \le \beta_0 - \delta_n\}$ converges to zero, with δ_n tuned as precisely as possible. Define a set $S_n^-(\beta) = S_n^-(\beta, S_n) = \{\beta' \in S_n : \beta' \le \beta\}$. The next theorem claims that the probability of underestimating β_0 for properly chosen δ_n converges to zero as $n \to \infty$.

Theorem 1. *Suppose $\theta \in \Theta_{\beta_0}$ for some $\beta_0 \in S_n$ and $|S_n^-(\beta_0)| \le C_1 \exp\{C_2 n^{1/(2\beta_0+1)}\}$ for some $C_1, C_2 > 0$. Then there exists an integer $N = N(\beta_0, \theta, C_2)$ such that for any $C > 0$ there exists a positive $C_0 = C_0(\beta_0, C_2, C)$ such that for all $c > C_0$ the inequality*

$$\mathrm{P}_\theta\Big\{\hat{\beta}(S_n) \le \beta_0 - \frac{c}{\log n}\Big\} < C_1 \exp\big\{-Cn^{1/(2\beta_0+1)}\big\}$$

holds for all $n \ge N$.

Remark 4. In fact $N = N(A_\theta(\beta_0), C_2)$. As it follows from the proof, the bigger $A_\theta(\beta_0)$, the bigger the corresponding $N(A_\theta(\beta_0), C_2)$. Even if $A_\theta(\beta) \uparrow \infty$ as $\beta \uparrow \bar{\beta}(\theta)$, by some subtle estimates one can show that there exists a sequence $\beta_n = \beta_n(\theta)$ such that $\beta_n \uparrow \bar{\beta}(\theta)$ (sufficiently slowly) and $\mathrm{P}_\theta\{\hat{\beta}(S_n) \le \beta_n\} \to 0$ as $n \to \infty$. The slower $\beta_n \uparrow \bar{\beta}(\theta)$, the faster $\mathrm{P}_\theta\{\hat{\beta}(S_n) \le \beta_n\} \to 0$ as $n \to \infty$.

Proof of Theorem 1. Notice that if $\beta_\theta(B) \in \mathcal{B}_0^\theta$, then $A_\theta(\beta_0) < \infty$. From Lemma 4 it follows that

$$\begin{aligned}\mathrm{P}_\theta\{\hat{\beta}(S_n) \le \beta_0 - \delta_n\} &\le \sum_{\beta \in S_n, \beta \le \beta_0 - \delta_n} \mathrm{P}_\theta\{\hat{\beta} = \beta\} \\ &\le C_1 \exp\Big\{\Big(C_2 + \frac{I(\beta_0)}{2}\Big) n^{1/(2\beta_0+1)} + \frac{5}{8} - \frac{n^{1/(2\beta_0+1-\delta_n)}}{16}\Big\}\end{aligned}$$

for all $n \ge N(\beta_0, \theta)$. Set $C' = C_2 + \frac{I(\beta_0)}{2}$. Then

$$\begin{aligned}&C' n^{1/(2\beta_0+1)} + \frac{5}{8} - \frac{1}{16} n^{1/(2\beta_0+1-\delta_n)} \\ &= \Big(C' + \frac{5}{8} n^{-1/(2\beta_0+1)} - \frac{1}{16} n^{\frac{\delta_n}{(2\beta_0+1-\delta_n)(2\beta_0+1)}}\Big) n^{1/(2\beta_0+1)} \\ &\le \Big(C' + 1 - \frac{n^{\delta_n/(2\beta_0+1)^2}}{16}\Big) n^{1/(2\beta_0+1)} \\ &= \Big(C' + 1 - \frac{1}{16} \exp\Big\{\frac{\delta_n}{(2\beta_0+1)^2} \log n\Big\}\Big) n^{1/(2\beta_0+1)} \\ &\le -Cn^{1/(2\beta_0+1)}\end{aligned}$$

for any $c > C_0$ with $C_0 = (2\beta_0+1)^2 \log(16(C' + 1 + C))$. The lemma follows for $C_0 = (2\beta_0+1)^2 \log\big[16(C_2 + 1 + C) + 8I(\beta_0)\big]$. □

Suppose that $\beta \geq \bar{\beta}(\theta)$, then it is reasonable to call $\mathrm{P}_\theta\big\{\hat{\beta}(S_n) \geq \beta\big\}$ the probability of oversmoothing. To claim good properties of $\hat{\beta}$ as some kind of smoothness estimator, we would like this probability to converge to zero. However, it turned out to be not the case without some extra condition. Intuitively, we can explain this by the fact that our statistics $\hat{\beta}$ is based on the prior which is designed for ℓ_2-risk based problem. So, the corresponding estimator $\hat{\theta}(\beta_2)$ for some $\beta > \bar{\beta}(\theta)$ can have a smaller ℓ_2-risk than $\hat{\theta}(\beta_1)$ for some $\beta_1 < \bar{\beta}(\theta)$. Therefore, our empirical Bayes procedure can pick some values $\beta \in S_n$, $\beta > \bar{\beta}(\theta)$, with significant nonzero probability. The condition below says essentially that $\theta \notin \Theta_\beta$ for all $\beta > \bar{\beta}(\theta)$ (or for all $\beta \geq \bar{\beta}(\theta)$ if $A_\theta(\bar{\beta}(\theta)) = \infty$) in more strict terms than just $A_\theta(\beta) = \infty$.

For a set $S \subseteq \mathbb{R}$, define $\lceil x \rceil_S = \inf\{y \in S : \ y \geq x\}$ and $\lfloor x \rfloor = \lceil x \rceil_{\mathbb{N}}$.

Condition A. Let $\bar{\beta}$ be the smoothness of θ. For $\beta \in S_n$, $\beta \geq \bar{\beta} + \delta_A$, $\delta_A > 0$ there exist $\beta' \equiv \beta'(\beta) \in S_n$, $\bar{\beta} < \beta' < \beta$, a constant $K_A > 0$, and an integer $N_A \equiv N_A(\bar{\beta})$ such that for $n \geq N_A$

$$\sum_{i=1}^{\infty} a_i(\beta, \beta')\Big(\frac{\theta_i^2}{2} - \tau_i^2(\beta')\Big) \geq (K_A - K_2(\bar{\beta}))n^{1-2\beta'/(2\beta+1)},$$

where $K_2(\bar{\beta}) = 2 + (2\bar{\beta})^{-1}$.

Define a set $V_\theta(\delta_A, K_A, N_A) = \{\beta \in S_n : \text{ Condition A is satisfied}\}$.

Example 1. If $\theta_i^2 = i^{-(2\bar{\beta}+1)}$ then for any $\delta_A > 0$, $\Delta > \delta_A$ there exists $N_A = N_A(\bar{\beta}, \delta_A, \Delta)$ such that Condition A is satisfied for $V_\theta = [\bar{\beta} + \delta_A, \bar{\beta} + \Delta]$ with $K_A = 2 + 3/(4\bar{\beta})$.

In the next theorem the probability of oversmoothing is established to converge to zero under Condition A.

Theorem 2. *Let $\bar{\beta}$ be the smoothness of θ. Let Condition A be fulfilled and the constants δ_A, K_A and N_A be defined in Condition A. Then there exists an integer $N = N(\theta, \delta_A, K_A, N_A)$ such that*

$$\mathrm{P}_\theta\big\{\widehat{\beta}(S_n) \geq \bar{\beta} + \delta_A\big\} \leq |S_n| \exp\Big\{-\frac{1}{2}\Big(C_A - K_2(\bar{\beta})\Big)n^{1/(2\bar{\beta}+3)}\Big\},$$

for all $n \geq N$, where $C_A = \min\{K_A, \frac{\delta_\varepsilon K}{8}\}$, $K_2(\bar{\beta}) = 1 + (2\bar{\beta})^{-1}$, $K > 8K_2(\bar{\beta})/(1 - 2^{-1/(2\bar{\beta}+5/2)})$.

Proof. The proof follows easily from Lemma 7 and remarks to this lemma. Indeed, we have from Lemma 7 and Remark 11 that for any $\beta \geq \bar{\beta} + \delta_A$ there exists $N = N(\theta, \delta_A, K_A, N_A)$ such that for all $n > N$

$$\mathrm{P}_\theta\{\widehat{\beta}(S_n) = \beta\} \leq \exp\Big\{-\frac{1}{2}\Big(C_A - K_2(\bar{\beta})\Big)n^{1/(2\bar{\beta}+3)}\Big\}.$$

Then

$$\mathrm{P}_\theta\big\{\widehat{\beta}(S_n) \geq \bar{\beta}+\delta_A\big\} = \sum_{\beta \in S_n \cap [\bar{\beta}+\delta_A, +\infty)} \mathrm{P}_\theta\big\{\widehat{\beta}(S_n) = \beta\big\} \leq |S_n| \exp\Big\{-\frac{1}{2}\Big(C_A - K_2(\bar{\beta})\Big)n^{1/(2\bar{\beta}+3)}\Big\}$$

for all $n > N$. $\square$

Remark 5. If $\delta_A > 1/2$ then Condition A is always satisfied (see Lemma 6). Thus there exists $N = N(\theta, K)$ such that for all $n \geq N$

$$\mathrm{P}_\theta\{\widehat{\beta}(S_n) > \bar{\beta} + \delta_A\} \leq |S_n| \exp\Big\{-\frac{1}{2}\Big(\frac{\delta_\varepsilon K}{8} - K_2(\bar{\beta})\Big) n^{1/(2\bar{\beta}+3)}\Big\}.$$

Remark 6. Let us give an intuitive explanation of Condition A. Let $\bar{\beta}$ be the smoothness of θ. Then $\theta \in \Theta_{\beta_0}$ for any $1/2 < \beta_0 < \bar{\beta}$. For each $\beta \in S_n$ consider the Bayesian estimator $\widehat{\theta}(\beta)$ of the parameter θ given by (3). For this estimator we can define bias and variance terms of the mean square risk. Denote them by $\mathrm{bias}_\theta(\beta)$ and $\mathrm{var}_\theta(\beta)$. Using Theorem 1 it is not difficult to show that for any β the variance term $\mathrm{var}_\theta(\beta)$ has the rate of convergence $n^{-2\beta_0/(2\beta_0+1)}$. The same is true for the bias term if $\beta < \bar{\beta}$.

Let $\beta > \bar{\beta}$. We believe that the following conjecture holds true. In particular, it implies that the plug-in estimator $\widehat{\theta}(\widehat{\beta})$ is an adaptive estimator of θ.

Conjecture. Let $\beta > \bar{\beta}$. If there exists $\beta' \in (\bar{\beta}, \beta)$ and a constant $d > 0$ such that $\sum_{i=1}^{\infty} a_i(\beta, \beta')(\theta_i^2/2 - \tau_i^2(\beta')) \geq dn^\gamma$, where $0 < \gamma \leq n^{1/(2\bar{\beta}+1)}$ for sufficiently large n, then

$$\mathrm{P}_\theta\{\widehat{\beta} = \beta\} \sim \exp\{-Cn^{1/(2\gamma)}\}, \quad n \to \infty.$$

Otherwise, for any $\beta_0 < \bar{\beta}$ there exist $C > 0$ and $N(\theta, \beta_0)$ such that for all $n \geq N$

$$\mathrm{bias}_\theta(\beta) \leq Cn^{-2\beta_0/(2\beta_0+1)}.$$

Shortly, Condition A controls the behavior of bias of the estimator $\widehat{\theta}(\beta)$.

Here we give an example of θ for which the conjecture holds true.

Example 2. Consider again $\theta \in \ell_2$ with components $\theta_i^2 = i^{-(2\bar{\beta}+1)}$. Let $\beta_0 > 1/2$ be fixed, $\beta_0 < \bar{\beta}$. Then certainly $\theta \in \Theta_{\beta_0}$. For any $\beta > \bar{\beta}$ consider the estimator $\widehat{\theta}(\beta)$. As it follows from Lemma 8, for any $\beta > \bar{\beta}$

$$\mathrm{bias}_\theta(\beta) = \sum_{i=1}^{\infty} \frac{i^{2(\beta-\bar{\beta})}}{(n + i^{2\beta+1})^2} \leq B(2\beta+1, 2, 2(\beta-\bar{\beta}))n^{-(2\bar{\beta}/(2\beta+1))}.$$

If $\beta < \bar{\beta}(2\beta_0+1)/(2\beta_0) - 1/2$ then $n^{-2\bar{\beta}/(2\beta+1)} < n^{-2\beta_0/(2\beta_0+1)}$. Thus we get $\mathrm{bias}_\theta(\beta) < B(\bar{\beta})n^{-2\beta_0/(2\beta_0+1)}$, where $B(\bar{\beta}) = \pi(2\bar{\beta}+1)^{-1}(\sin(\pi/(2\bar{\beta}+1)))^{-1}$ upper bounds the constant $B(2\beta+1, 2, 2(\beta-\bar{\beta}))$ (see Lemma 9).

Now, if $\beta > \bar{\beta} + 1$ then Condition A is satisfied as it is shown in the second part of Lemma 6. If $\bar{\beta}+1 \geq \beta > \bar{\beta}(2\beta_0+1)/(2\beta_0) - 1/2$, then we have from Lemmas 8 and 9 that there exists an integer $N(\beta_0, \bar{\beta})$ such that for any $n \geq N$, for any $\beta > \beta' > \bar{\beta}$

$$\begin{aligned}\sum_{i=1}^{\infty} a_i(\beta, \beta')\Big(\frac{\theta_i^2}{2} - \tau_i^2(\beta')\Big) &= \frac{1}{2}\sum_{i=1}^{\infty}\frac{ni^{2(\beta-\bar{\beta})}}{n + i^{2\beta+1}} - \frac{1}{2}\sum_{i=1}^{\infty}\frac{ni^{2(\beta'-\bar{\beta})}}{n + i^{2\beta+1}} \\ &\quad - \sum_{i=1}^{\infty}\frac{ni^{2(\beta-\beta')}}{n + i^{2\beta+1}} + \sum_{i=1}^{\infty}\frac{n}{n + i^{2\beta'+1}} \\ &\geq 1/(4\bar{\beta})n^{1-2\bar{\beta}/(2\beta+1)} > 1/(4\bar{\beta})n^{1/(2\bar{\beta}+1)}.\end{aligned}$$

Thus Condition A is satisfied for any $\beta > \bar{\beta}(2\beta_0+1)/(2\beta_0) - 1/2$. Then from Lemma 7 it follows that $\mathrm{P}_\theta\{\widehat{\beta} = \beta\}$ decreases exponentially as $n \to \infty$.

4 Auxiliary results

This section provides some auxiliary lemmas which we need to prove the main results.

Lemma 3. *Suppose $|S_n| < \infty$. Then for any $\beta_0, \beta \in S_n$*

$$\mathrm{P}_\theta\big\{\hat{\beta}(S_n) = \beta\big\} \le \exp\Big\{\frac{1}{2}\sum_{i=1}^\infty \frac{(\tau_i^2(\beta) - \tau_i^2(\beta_0))(\theta_i^2 - \tau_i^2(\beta_0))}{\tau_i^2(\beta)\tau_i^2(\beta_0) + 2n^{-1}\tau_i^2(\beta_0) + n^{-2}}\Big\}.$$

Proof. We use here the following shorthand notations: $a_i = a_i(\beta, \beta_0)$, $b_i = b_i(\beta, \beta_0)$.

Since $\beta_0 \in S_n$, by (8) and the Markov inequality, we have

$$\begin{aligned}
\mathrm{P}_\theta\{\hat{\beta} = \beta\} &= \mathrm{P}_\theta\big\{Z_n(\beta, \beta') \le 0 \ \forall\, \beta' \in S_n\big\} \le \mathrm{P}_\theta\big\{Z_n(\beta,\beta_0) \le 0\big\} \\
&= \mathrm{P}_\theta\Big\{-\sum_{i=1}^\infty a_i X_i^2 \ge \sum_{i=1}^\infty \log b_i\Big\} \\
&\le \mathrm{E}_\theta \exp\Big\{-\frac{1}{2}\sum_{i=1}^\infty a_i X_i^2\Big\} \exp\Big\{-\frac{1}{2}\sum_{i=1}^\infty \log b_i\Big\}.
\end{aligned}$$

To compute $\mathrm{E}_\theta \exp\{-\frac{1}{2}a_i X_i^2\}$, we use the following elementary identity for a Gaussian random variable $\eta \sim \mathcal{N}(\mu, \sigma^2)$:

$$\mathrm{E}\exp\{\kappa\eta^2\} = (1-2\kappa\sigma^2)^{-1/2}\exp\Big\{\frac{\kappa\mu^2}{1-2\kappa\sigma^2}\Big\}, \qquad \text{for} \quad \kappa < \frac{1}{2\sigma^2}.$$

Apply this equality for $\kappa = -\frac{a_i}{2}$ and $\eta = X_i$ (condition $\kappa < \frac{1}{2\sigma^2}$ corresponds to $-a_i < n$ which is always true since $|a_i| < n$ for all $i \in \mathbb{N}$):

$$\mathrm{E}_\theta \exp\Big\{-\frac{1}{2}a_i X_i^2\Big\} = (1+n^{-1}a_i)^{-1/2}\exp\Big\{\frac{-a_i\theta_i^2}{2(1+n^{-1}a_i)}\Big\}.$$

Combining the previous relations, we obtain

$$\mathrm{P}_\theta\{\hat{\beta} = \beta\} \le \prod_{i=1}^\infty \frac{b_i^{-1/2}}{(1+n^{-1}a_i)^{1/2}}\exp\Big\{\frac{-a_i\theta_i^2}{2(1+n^{-1}a_i)}\Big\}. \tag{9}$$

From the definitions (6) and (7) it follows

$$\frac{b_i^{-1}}{1+n^{-1}a_i} = 1 + \frac{a_i\tau_i^2(\beta_0)}{1+n^{-1}a_i}.$$

Using this, the elementary inequality $1 + x \le e^x$, $x \in \mathbb{R}$, and (9), we finally arrive at

$$\mathrm{P}_\theta\{\hat{\beta} = \beta\} \le \exp\Big\{\frac{1}{2}\sum_{i=1}^\infty \frac{a_i\big(\tau_i^2(\beta_0) - \theta_i^2\big)}{1+n^{-1}a_i}\Big\} = \exp\Big\{\frac{1}{2}\sum_{i=1}^\infty \frac{(\tau_i^2(\beta) - \tau_i^2(\beta_0))(\theta_i^2 - \tau_i^2(\beta_0))}{\tau_i^2(\beta)\tau_i^2(\beta_0) + 2n^{-1}\tau_i^2(\beta_0) + n^{-2}}\Big\}.$$

□

The proof of Theorem 1 is based on the following lemma.

Lemma 4. *Let $A_\theta(\beta_0) < \infty$ for some $\beta_0 \in S_n$, $|S_n| < \infty$. Then for any $\beta \in S_n$, $\beta < \beta_0$, there exists $N = N(\beta_0, \theta)$ such that for any $n \geq N$*

$$\mathrm{P}_\theta\big\{\hat\beta(S_n) = \beta\big\} \leq \exp\Big\{\frac{I(\beta_0) n^{1/(2\beta_0+1)}}{2} + \frac{5}{8} - \frac{n^{1/(\beta+\beta_0+1)}}{16}\Big\},$$

where $I(\beta_0) = B(2\beta_0 + 1, 2, 0)$ is defined by (4).

Proof. We make use of Lemma 3:

$$\text{(10)} \quad \mathrm{P}_\theta\{\hat\beta = \beta\} \leq \exp\Big\{\frac{1}{2}\sum_{i=1}^\infty \frac{(\tau_i^2(\beta) - \tau_i^2(\beta_0))(\theta_i^2 - \tau_i^2(\beta_0))}{\tau_i^2(\beta)\tau_i^2(\beta_0) + 2n^{-1}\tau_i^2(\beta_0) + n^{-2}}\Big\} = \exp\Big\{\frac{S_1 + S_2(\theta)}{2}\Big\},$$

where

$$S_1 = \sum_{i=1}^\infty \frac{\tau_i^4(\beta_0) - \tau_i^2(\beta)\tau_i^2(\beta_0)}{\tau_i^2(\beta)\tau_i^2(\beta_0) + 2n^{-1}\tau_i^2(\beta_0) + n^{-2}} = S_{11} - S_{12}\,,$$

$$S_2(\theta) = \sum_{i=1}^\infty \frac{-a_i(\beta,\beta_0)\theta_i^2}{1 + n^{-1}a_i(\beta,\beta_0)} = \sum_{i=1}^\infty \frac{\big(\tau_i^2(\beta) - \tau_i^2(\beta_0)\big)\theta_i^2}{\tau_i^2(\beta)\tau_i^2(\beta_0) + 2n^{-1}\tau_i^2(\beta_0) + n^{-2}}\,.$$

The rest of the proof is very much the same as the proof of Lemma 3.1 in Belitser and Ghosal (2003). First we bound term S_1. As $\beta < \beta_0$, $i^{-(2\beta+1)} > i^{-(2\beta_0+1)}$ and therefore, by Lemma 8, we obtain

$$\begin{aligned} S_{11} &= \sum_{i=1}^\infty \frac{i^{-2(2\beta_0+1)}}{i^{-2(\beta+\beta_0+1)} + 2n^{-1}i^{-(2\beta_0+1)} + n^{-2}} \leq \sum_{i=1}^\infty \frac{i^{-2(2\beta_0+1)}}{(i^{-(2\beta_0+1)} + n^{-1})^2} \\ &= n^2 \sum_{i=1}^\infty \frac{1}{(n + i^{2\beta_0+1})^2} \leq B(2\beta_0 + 1, 2, 0) n^{1/(2\beta_0+1)} + 1\,. \end{aligned}$$

To bound S_{12} from below, note first that the term $i^{-2(\beta+\beta_0+1)}$ is not less than n^{-2} for $i \leq n^{1/(\beta+\beta_0+1)}$ and not less than $n^{-1}i^{-(2\beta_0+1)}$ for $i \leq n^{1/(2\beta+1)}$ (which includes all $i \leq n^{1/(\beta+\beta_0+1)}$ since $\beta < \beta_0$). This implies

$$S_{12} = \sum_{i=1}^\infty \frac{i^{-2(\beta_0+\beta+1)}}{i^{-2(\beta+\beta_0+1)} + 2n^{-1}i^{-(2\beta_0+1)} + n^{-2}} \geq \sum_{i=1}^{\lfloor n^{1/(\beta+\beta_0+1)}\rfloor} \frac{i^{-2(\beta+\beta_0+1)}}{4i^{-2(\beta+\beta_0+1)}} \geq \frac{\lfloor n^{1/(\beta+\beta_0+1)}\rfloor}{4}\,.$$

Combining the last two inequalities, we arrive at

$$\text{(11)} \qquad S_1 \leq B(2\beta_0 + 1, 2, 0) n^{1/(2\beta_0+1)} - \frac{n^{1/(\beta+\beta_0+1)}}{4} + \frac{5}{4}\,.$$

Now note that $\tau_i^2(\beta) > \tau_i^2(\beta_0)$ as $\beta < \beta_0$. Then, for some integer M, we have

$$\begin{aligned} S_2(\theta) &\leq \sum_{i=1}^\infty \frac{\tau_i^2(\beta)\theta_i^2}{\tau_i^2(\beta)\tau_i^2(\beta_0) + 2n^{-1}\tau_i^2(\beta_0) + n^{-2}} = \sum_{i=1}^\infty \frac{n^2 i^{2\beta_0+1}\theta_i^2}{n^2 + 2ni^{2\beta+1} + i^{2(\beta+\beta_0+1)}} \\ &\leq \sum_{i=1}^M i^{2\beta_0+1}\theta_i^2 + \sum_{i=M+1}^\infty \frac{n^2 i^{2\beta_0}\theta_i^2}{i^{2\beta+2\beta_0+1}} \leq M\sum_{i=1}^M i^{2\beta_0}\theta_i^2 + \frac{n^2}{M^{2\beta+2\beta_0+1}} \sum_{i=M+1}^\infty i^{2\beta_0}\theta_i^2\,. \end{aligned}$$

Take $M = M_n = M_n(\beta_0, \beta, A_\theta(\beta_0)) = \lfloor \varepsilon(2A_\theta(\beta_0))^{-1} n^{1/(\beta+\beta_0+1)} \rfloor + 1$ for some $\varepsilon > 0$ so that $M \sum_{i=1}^{M} i^{2\beta_0}\theta_i^2 \le \varepsilon/2$ for all $n \ge N_1 = N_1(\beta_0)$. Since $A_\theta(\beta_0) < \infty$, there exists $N_2 = N_2(\beta_0, \theta, \varepsilon)$ such that for any $n \ge N_2$

$$\sum_{i=M_n+1}^{\infty} i^{2\beta_0}\theta_i^2 \le \frac{\varepsilon}{2}\big(2A_\theta(\beta_0)/\varepsilon\big)^{-4\beta_0-1} \le \frac{\varepsilon}{2}\big(2A_\theta(\beta_0)/\varepsilon\big)^{-2\beta-2\beta_0-1}.$$

which implies that the second term is also at most $\varepsilon/2$ for all $n \ge N_2$. Therefore $S_2(\theta) \le \varepsilon n^{1/(\beta+\beta_0+1)}$ for all $n \ge \max\{N_1, N_2\}$. We choose $\varepsilon = 1/8$ and combine the last relation with (10) and (11) to finish the proof. □

Remark 7. The above result is not uniform with respect to θ since the inequality holds only for $n \ge N(\beta_0, \theta)$. In general, the result is not uniform over the the set $\theta \in \Theta_{\beta_0}(\bar\theta, Q) = \{\theta : A_{\theta-\bar\theta}(\beta_0) \le Q\}$, which is an ellipsoid of "size" Q around $\bar\theta$. However, if $A_{\bar\theta}(\beta_0) < \infty$, then for a sufficiently small ellipsoid size Q, the uniformity does hold. Indeed, we need only to evaluate the term $S_2(\theta)$. Now, for any $\theta \in \Theta_{\beta_0}(\bar\theta, Q)$ we have $S_2(\theta) \le 2S_2(\bar\theta) + 2S_2(\theta - \bar\theta)$. As in the proof of Lemma 4, we can find $N_1 = N_1(\beta_0, \bar\theta, \varepsilon)$ such that $S_2(\bar\theta) \le \varepsilon/4$ for all $n \ge N_1$. Next, by taking $M = M_n = \lfloor (n^{1/(\beta+\beta_0+1)} \rfloor + 1$, we derive that for any $Q < \varepsilon/4$ there exists $N_2 = N_2(\beta_0, Q)$ such that $S_2(\theta - \bar\theta) \le A_{\theta-\bar\theta}(\beta_0) n^{1/(\beta+\beta_0+1)}\big(1 + n^{-(\beta+\beta_0+1)}\big) \le \varepsilon n^{1/(\beta+\beta_0+1)}/4$ for all $n \ge N_2$ for any $\theta \in \Theta_{\beta_0}(\bar\theta, Q)$. We conclude that for any $Q < \varepsilon/4$ there exists $N_3 = N_3(\beta_0, \bar\theta, Q, \varepsilon) = \max\{N_1(\beta_0, \bar\theta, \varepsilon), N_2(\beta_0, Q)\}$ such that $S_2(\theta) \le \varepsilon n^{1/(\beta+\beta_0+1)}$ for all $n \ge N_3$, uniformly over $\theta \in \Theta_{\beta_0}(\bar\theta, Q)$. Take $\varepsilon = 1/8$ to derive the assertion of the lemma uniformly over $\theta \in \Theta_{\beta_0}(\bar\theta, Q)$ for any $Q < 1/32$.

Below are some auxiliary lemmas, which we will use in the proof of Theorem 12.

Lemma 5. *For any $\beta_0, \beta', \beta \in \mathcal{B}$ such that $\beta_0 \le \beta' < \beta$ the following inequality holds:*

$$\sum_{i=1}^{\infty} \frac{\tau_i^2(\beta')}{\tau_i^2(\beta) + n^{-1}} \le K_2 n^{1-2\beta'/(2\beta+1)},$$

where $K_2 = K_2(\beta_0) = 2 + (2\beta_0)^{-1}$.

Proof. Notice that

$$\sum_{i=1}^{\infty} \frac{\tau_i^2(\beta')}{\tau_i^2(\beta) + n^{-1}} = \sum_{i=1}^{\infty} \frac{n i^{2\beta-2\beta'}}{n + i^{2\beta+1}}$$

The sum in the right hand side of the above inequality is bounded from above by

$$\sum_{i=1}^{\lfloor n^{1/(2\beta+1)} \rfloor} i^{2\beta-2\beta'} + \sum_{i=\lfloor n^{1/(2\beta+1)} \rfloor+1}^{\infty} n i^{-(2\beta'+1)}$$

$$\le \int_{1}^{\lfloor n^{1/(2\beta+1)} \rfloor} x^{2\beta-2\beta'}\,dx + \lfloor n^{1/(2\beta+1)} \rfloor + \int_{\lfloor n^{1/(2\beta+1)} \rfloor+1}^{\infty} n x^{-(2\beta'+1)}\,dx$$

$$< \left(\frac{1}{2\beta - 2\beta' + 1} + \frac{1}{2\beta'} + 1\right) n^{1-2\beta'/(2\beta+1)}.$$

Combining this with $\beta_0 \le \beta' < \beta$ completes the proof of the lemma. $\square$

Remark 8. We can also obtain exact constant in the above upper bound using Lemma 9.

Then next lemma shows that Condition A is satisfied for $\beta \in S_n$, $\beta \ge \bar\beta + 1/2 + \varepsilon$ for any $\varepsilon > 0$.

Lemma 6. *Let $\beta \in [\bar\beta + 1/2 + \varepsilon, +\infty) \cap S_n$ for some $\varepsilon > 0$ and $K_2(\bar\beta) = 2 + (2\bar\beta)^{-1}$, where $\bar\beta$ is the smoothness of θ.*

1. *If $0 < \varepsilon \le 1/2$ then there exist $\beta' \in (\bar\beta + 1/2 + \varepsilon, \beta)$ and an integer $N = N(\theta, K, \varepsilon)$ such that for all $n \ge N$*
$$\sum_{i=1}^{\infty} a_i(\beta,\beta')\Big(\frac{\theta_i^2}{2} - \tau_i^2(\beta')\Big) \ge \Big(\frac{\delta_\varepsilon K}{8} - K_2(\bar\beta)\Big) n^{1/(2\bar\beta+2+\varepsilon)},$$
where $\delta_\varepsilon = 1 - 2^{-1/(2\bar\beta+2+\varepsilon)}$, $K > 8K_2(\bar\beta)/\delta_\varepsilon$.

2. *If $\varepsilon > 1/2$ then there exists $\beta' \in (\bar\beta + 1, \beta)$ and an integer $N = N(\theta, K)$ such that for all $n \ge N$*
$$\sum_{i=1}^{\infty} a_i(\beta,\beta')\Big(\frac{\theta_i^2}{2} - \tau_i^2(\beta')\Big) \ge \Big(\frac{\delta K}{8} - K_2(\bar\beta)\Big) n^{1/(2\bar\beta+1)},$$
where $\delta = 1 - 4^{-1/(2\bar\beta+1)}$, $K > 8K_2(\bar\beta)/\delta$.

Proof. First, notice that from Lemma 5 we get that for any $\beta' \in (\bar\beta, \beta)$

$$\sum_{i=1}^{\infty} a_i(\beta,\beta')\tau_i^2(\beta') < \sum_{i=1}^{\infty} \frac{\tau_i^2(\beta')}{\tau_i^2(\beta) + n^{-1}} \le K_2(\bar\beta) n^{1-2\beta'/(2\beta+1)}. \tag{12}$$

Next, for any $\beta' < \beta$, $0 \le \delta \le 1 - \exp\{-2(\log 2)(\beta - \beta')\}$ and $T_n = \lfloor n^{1/(2\beta+1)} \rfloor$ we have

$$\sum_{i=1}^{\infty} a_i(\beta,\beta')\theta_i^2 \ge \delta \sum_{i=2}^{\infty} \frac{n^2 i^{2\beta+1}\theta_i^2}{(i^{2\beta+1}+n)(i^{2\beta'+1}+n)} \ge \frac{\delta}{2}\sum_{i=2}^{T_n} \frac{n i^{2\beta+1}\theta_i^2}{i^{2\beta+1}+n} \ge \frac{\delta}{4}\sum_{i=2}^{T_n} i^{2\beta+1}\theta_i^2. \tag{13}$$

Let $\beta \ge \bar\beta + 1/2 + \varepsilon$. Since $\sum_{i=1}^{\infty} i^{2\bar\beta+\varepsilon/2}\theta_i^2 = \infty$ for any $\varepsilon > 0$, there exist infinitely many $i \in \mathbb{N}$ such that $i^{2\bar\beta+\varepsilon/2}\theta_i^2 \ge K i^{-1-\varepsilon/2}$ for any $K > 0$. Thus for $\beta > \bar\beta + 1/2$ we have

$$\frac{\delta}{4}\sum_{i=2}^{T_n} i^{2\beta+1}\theta_i^2 = \frac{\delta}{4}\sum_{i=2}^{T_n} i^{2\beta-2\bar\beta+1-\varepsilon/2}\theta_i^2 i^{2\bar\beta+\varepsilon/2} \ge \frac{\delta}{4}\frac{K}{T_n^{1+\varepsilon/2}} T_n^{2\beta-2\bar\beta+1-\varepsilon/2} \ge \frac{\delta K}{4} n^{(2\beta-2\bar\beta+\varepsilon)/(2\beta+1)}$$

for infinitely many n. Certainly $n^{(2\beta-2\bar\beta+\varepsilon)/(2\beta+1)} \ge n^{1-2\beta'/(2\beta+1)}$ for any $\beta' \ge \bar\beta + \frac{1+\varepsilon}{2}$. Thus for any $\beta > \bar\beta + 1/2 + \varepsilon/2$, there exists $\beta' \in (\bar\beta + 1/2 + \varepsilon, \beta)$ such that

$$\sum_{i=1}^{\infty} a_i(\beta,\beta')\theta_i^2 \ge \frac{\delta}{2}\sum_{i=2}^{T_n} \frac{n i^{2\beta+1}\theta_i^2}{i^{2\beta+1}+n} \ge \frac{\delta}{4}\sum_{i=2}^{T_n} i^{2\beta+1}\theta_i^2 \ge \frac{\delta K}{4} n^{1-2\beta'/(2\beta+1)} \tag{14}$$

for any $K > 0$ and infinitely many n. This inequality holds not just for infinitely many n, but also for all sufficiently large n. This is true by the following arguments: it is easy to show that, for any $\alpha \in (0,1)$, $n^{-\alpha}\sum_{i=2}^{T_n} \frac{ni^{2\beta+1}\theta_i^2}{i^{2\beta+1}+n}$ is a decreasing function of n. Therefore if this function is bigger than a certain constant for infinitely many n, then this inequality must hold for all $n \geq N_0$ for some $N_0 = N_0(\theta, K)$.

Combining estimates (12) and (14) we get that for $\beta \geq \bar\beta + 1/2 + \varepsilon$ there exists β' such that for sufficiently large n

$$\sum_{i=1}^{\infty} a_i(\beta,\beta')\Big(\frac{\theta_i^2}{2} - \tau_i^2(\beta')\Big) \geq \Big(\frac{\delta K}{8} - K_2(\bar\beta)\Big) n^{1-2\beta'/(2\beta+1)}.$$

Notice that if we choose $\beta' = [\bar\beta(2\beta+1)/(2\bar\beta+1)]_{S_n}$ then $n^{1-2\beta'/(2\beta+1)} \geq n^{1/(2\bar\beta+1)}$ and at the same time $\bar\beta < \beta' < \beta$. This choice is possible only for $\beta > \bar\beta + 1$, since only in this case $\beta' > \bar\beta + 1/2$. In this case $\beta - \beta' > 1/(2\bar\beta+1)$, thus taking $\delta = 1 - \exp\{-2\log 2/(2\bar\beta+1)\}$ we get from (12) and (13) that for $\beta > \bar\beta + 1$ there exists $\beta' \in (\bar\beta+1, \beta)$ such that

$$\sum_{i=1}^{\infty} a_i(\beta,\beta')\Big(\frac{\theta_i^2}{2} - \tau_i^2(\beta')\Big) \geq \Big(\frac{\delta K}{8} - K_2(\bar\beta)\Big) n^{1/(2\bar\beta+1)}.$$

If $\beta \in [\bar\beta + 1/2 + \varepsilon, +\infty)$ then choose $\beta' = [(\bar\beta + 1/2 + \varepsilon/2)(2\beta+1)(2\bar\beta+2+\varepsilon)]_{S_n}$. In this case $\beta - \beta' > \varepsilon/(2(2\bar\beta+2+\varepsilon))$ and $n^{1-2\beta'/(2\beta+1)} \geq n^{1/(2\bar\beta+2+\varepsilon)}$. Thus, taking $\delta = \delta_\varepsilon = 1 - \exp\{-\log 2\varepsilon/(2\bar\beta+2+\varepsilon)\}$ we get analogously that for all $n \geq N(\theta,\varepsilon,K)$

$$\sum_{i=1}^{\infty} a_i(\beta,\beta')\Big(\frac{\theta_i^2}{2} - \tau_i^2(\beta')\Big) \geq \Big(\frac{\delta_\varepsilon K}{8} - K_2(\bar\beta)\Big) n^{1/(2\bar\beta+2+\varepsilon)}.$$

The lemma is proved. □

Remark 9. If Condition A is satisfied then for $\beta \in (\bar\beta + \delta_A, \bar\beta + 1]$ we have for all $n \geq N_A$

$$\sum_{i=1}^{\infty} a_i(\beta,\beta')\Big(\frac{\theta_i^2}{2} - \tau_i^2(\beta')\Big) \geq \Big(C_A - K_2(\bar\beta)\Big) n^{1/(2\bar\beta+3)},$$

where $C_A = \min\{K_A, \delta_\varepsilon K/8\}$. This fact follows from the following relation: $n^{1-2\beta'/(2\beta+1)} > n^{1/(2\beta'+1)} > n^{1/(2\beta+1)} > n^{1/(2\bar\beta+3)}$.

Lemma 7. *Let $\bar\beta$ be the smoothness of θ, $\beta \in V_\theta(\delta_A, K_A, N_A)$ and β' be chosen as in Condition A. Then for all $n \geq N_A$*

$$\mathrm{P}_\theta\{\widehat\beta = \beta\} \leq \exp\Big\{-\frac{1}{2}\Big(K_A - K_2(\bar\beta)\Big) n^{1-2\beta'/(2\beta+1)}\Big\}.$$

Proof. For any $\beta' < \beta$ we have $a_i(\beta,\beta') > 0$. For β' chosen to satisfy Condition A we obtain from Lemma 3

$$\mathrm{P}_\theta\{\hat\beta = \beta\} \leq \exp\Big\{\frac{1}{2}\sum_{i=1}^{\infty} \frac{a_i(\beta,\beta')\big(\tau_i^2(\beta') - \theta_i^2\big)}{1 + n^{-1}a_i(\beta,\beta')}\Big\}.$$

Since $0 < a_i(\beta,\beta') < n$, we have

$$\sum_{i=1}^{\infty} \frac{a_i(\beta,\beta')\big(\tau_i^2(\beta')-\theta_i^2\big)}{1+n^{-1}a_i(\beta,\beta')} < \sum_{i=1}^{\infty} a_i(\beta,\beta')\Big(\tau_i^2(\beta')-\frac{\theta_i^2}{2}\Big).$$

We upper bound the right-hand side using Condition A and prove the lemma. □

Remark 10. For $\beta > \bar\beta + 1/2$ the probability $\mathrm{P}_\theta\{\widehat\beta = \beta\}$ decreases exponentially even if Condition A is not satisfied. Namely, let $\beta \in [\bar\beta + 1/2 + \varepsilon, +\infty) \cap S_n$ for some $\varepsilon > 0$.

1. If $0 < \varepsilon \le 1/2$, then there exists $N = N(\theta, K_A)$ such that for all $n \ge N$

$$\mathrm{P}_\theta\{\widehat\beta = \beta\} \le \exp\Big\{-\frac{1}{2}\Big(\frac{\delta_\varepsilon K}{8} - K_2(\bar\beta)\Big)n^{1/(2\bar\beta+2+\varepsilon)}\Big\}.$$

2. If $\varepsilon > 1/2$, for all $n \ge N$

$$\mathrm{P}_\theta\{\widehat\beta = \beta\} \le \exp\Big\{-\frac{1}{2}\Big(\frac{\delta K}{8} - K_2(\bar\beta)n^{1/(2\bar\beta+1)}\Big\}.$$

The proof is similar to the proof of Lemma 7 and follows immediately from Lemma 6.

Remark 11. For $\beta > \bar\beta + \delta_A$ there exists an integer $N = N(\theta, \delta_A, K_A, N_A)$ such that for all $n \ge N$

$$\mathrm{P}_\theta\{\widehat\beta = \beta\} \le \exp\Big\{-\frac{1}{2}\Big(C_A - K_2(\bar\beta)\Big)n^{1/(2\bar\beta+3)}\Big\}$$

where $C_A = \min\{K_A, \delta K/8\}$, $K > 8K_2(\bar\beta)/(1-2^{-1/(2\bar\beta+5/2)})$.

This result follows immediately from Remarks 9, 10 and Lemma 7 if we compare the constants and rates in the upper bounds taking into account that $\delta_\varepsilon < 1/(1-2^{-1/(2\bar\beta+5/2)})$ and $\delta_\varepsilon < \delta$.

5 Goodness-of-fit testing

In this section we discuss some implications of Theorems 1 and 2 related to hypotheses testing. We can use the statistic $\widehat\beta(X)$ from (5) in the following goodness-of-fit problem.

We observe independent Gaussian data $X = (X_i)_{i=1}^\infty$ such that $X_i \sim \mathcal{N}(\theta_i, n^{-1})$ where $\theta \in \Theta_\infty$. Here $\Theta_\infty = \cup_{\beta\in\mathcal{B}}\Theta_\beta$. Let $\mathcal{P} = \{P_{\theta,n},\ \theta \in \Theta_\infty\}$ be a set of Gaussian probability measures with mean θ and covariance matrix $n^{-1}I$. Fix some $\beta_0 \in \mathcal{B}$. We would like to test the hypothesis $H_0:\ \theta \in \Theta_{\beta_0}$. It would be ideal to test this hypothesis against the alternative $H_1:\ \theta \in \Theta_\infty \setminus \Theta_{\beta_0}$. Unfortunately, as it was realized by many researchers, this problem has no solution. Typical approach in this case is to restrict the set of alternatives (see [11]).

Define a set

$$\Lambda_{\beta_0,n}^{\delta_A} = \{\theta:\ \sum_{i=1}^{\infty}\theta_i^2 i^{2(\beta_0-\delta_n)} = \infty, \quad V_\theta(\delta_A, K_A, N_A) \supseteq (\beta_0+\delta_A, \infty)\},$$

where $K_A > K_2(\beta_0)$, δ_A and N_A are defined by Condition A, δ_n is a positive sequence. Evidently, $\Lambda^{\delta_A}_{\beta_0,n} \cap \Theta_{\beta_0} = \emptyset$, but $(\Theta_\infty \setminus \Lambda^{\delta_A}_{\beta_0,n}) \cap \Theta_{\beta_0} \neq \emptyset$. In fact, $\Lambda^{\delta_A}_{\beta_0,n}$ is a set of all θ's for which Condition A is satisfied for all $\beta > \beta_0 + \delta_A$.

We are going to test the hypothesis

$$H_0: \ \theta \in \Theta_{\beta_0+\delta_A} \text{ against } H_1: \ \theta \in \Lambda^{\delta_A}_{\beta_0,n}.$$

The decision rule is a function

$$\psi_n(X) = \mathbf{1}\{\widehat{\beta}(X) < \beta_0 - \delta_n\}$$

where $\widehat{\beta}(X)$ is the marginal maximum likelihood estimator (5).

Introduce two disjoint subsets of $\mathcal{P}$ corresponding to our hypotheses:

$$\mathcal{P}_0 = \{P_{\theta,n}: \ \theta \in \Theta_{\beta_0+\delta_A}\}, \quad \mathcal{P}_1 = \{P_{\theta,n}: \ \theta \in \Lambda^{\delta_A}_{\beta_0,n}\}.$$

The probabilities of type I and II errors

$$\alpha_1(\psi) \equiv \alpha_1(\psi, P) = \mathrm{E}_P\psi, \ P \in \mathcal{P}_0; \qquad \alpha_2(\psi) \equiv \alpha_2(\psi, P) = \mathrm{E}_P(1-\psi), \ P \in \mathcal{P}_1$$

can be written in equivalent form

$$\alpha_1(\psi,\theta) = \mathrm{P}_\theta\{\theta \in \Lambda^{\delta_A}_{\beta_0,n}\}, \ \theta \in \Theta_{\beta_0+\delta_A}; \qquad \alpha_2(\psi,\theta) = \mathrm{P}_\theta\{\theta \in \Theta_{\beta_0+\delta_A}\}, \ \theta \in \Lambda^{\delta_A}_{\beta_0,n}.$$

Notice that $\{\theta \in \Lambda^{\delta_A}_{\beta_0,n}\} \subset \{\theta \notin \Theta_{\beta_0-\delta_n}\}$. Consequently, we have

$$\alpha_1(\psi,\theta) \le \mathrm{P}_\theta\{\widehat{\beta} < \beta_0 - \delta_n\}, \ \theta \in \Theta_{\beta_0+\delta_A}; \qquad \alpha_2(\psi,\theta) = \mathrm{P}_\theta\{\widehat{\beta} > \beta_0 + \delta_A\}, \ \theta \in \Lambda^{\delta_A}_{\beta_0,n}.$$

Theorem 3. *Set $\beta_0 \in \mathcal{B}$.*

1. *Let $|S_n^-(\beta_0)| \le C_1 \exp\{C_2 n^{1/(2\beta_0+1)}\}$ for some $C_1, C_2 > 0$. For any $\theta \in \Theta_{\beta_0}$ there exists an integer $N = N(\beta_0,\theta,C_2)$ such that for any $C > 0$ there exists a positive $C_0 = C_0(\beta_0, C_2, C)$ such that for all $c > C_0$ and for all $n > N$*

$$\alpha_1(\psi,\theta) \le \mathrm{P}_\theta\{\widehat{\beta} < \beta_0 - \frac{c}{\log n}\} < C_1 \exp\{-Cn^{1/(2\beta_0+1)}\}$$

2. *Let $\theta \in \Lambda^{\delta_A}_{\beta_0,n}$. Then there exists an integer $N = N(\theta, \beta_A, K_A, N_A)$ such that*

$$\alpha_2(\psi,\theta) \equiv \mathrm{P}_\theta\{\widehat{\beta} > \beta_0 + \delta_A\} \le |S_n| \exp\Big\{-\frac{1}{2}\Big(C_A - K_2(\beta_0 - \delta_n)\Big) n^{1/(2\beta_0+3-2\delta_n)}\Big\},$$

for all $n \ge N$, where the constant C_A is defined in Theorem 2 and $K_2(\beta_0 - \delta_n) = 2 + 1/(2\beta_0 - 2\delta_n)$.

Proof. The proof of the first part immediately follows from Theorem 1. As to the second part, notice that if $\theta \in \Lambda^{\delta_A}_{\beta_0,n}$ then $\bar{\beta} \le \beta_0 - \delta_n$ and

$$\mathrm{P}_\theta\{\widehat{\beta} > \beta_0 + \delta_A\} \le \mathrm{P}_\theta\{\widehat{\beta} > \bar{\beta} + \delta_n + \delta_A\} \le \{\widehat{\beta} > \bar{\beta} + \delta_A\}.$$

Applying now Theorem 2 we compete the proof. □

Remark 12. If $\delta_A > 1/2$ then we can construct consistent test for the problem of testing

$$H_0: \ \theta \in \Theta_{\beta_0+\delta_A} \quad \text{against } H_1: \ \theta \notin \Theta_{\beta_0-\delta_n}.$$

In this case for the same decision rule the probabilities of errors have the form

$$\alpha_1(\psi,\theta) = \mathrm{P}_\theta\{\theta \notin \Theta_{\beta_0-\delta_n}\} = \mathrm{P}_\theta\{\widehat{\beta} < \beta_0 - \delta_n\}, \ \theta \in \Theta_{\beta_0+\delta_A}$$
$$\alpha_2(\psi,\theta) = \mathrm{P}_\theta\{\theta \in \Theta_{\beta_0+\delta_A}\} = \mathrm{P}_\theta\{\widehat{\beta} > \beta_0 + \delta_A\}, \ \theta \notin \Theta_{\beta_0-\delta_n}.$$

Then $\alpha_2(\psi,\theta) < \mathrm{P}_\theta\{\widehat{\beta} > \bar{\beta}+\delta_n+1/2\}$ and from Remark 5 we get that our test is consistent and asymptotically unbiased. Namely, $\alpha_1(\psi,\theta) \to 0$ and $\alpha_2(\psi,\theta) \to 0$ as $n \to \infty$.

6 Discussion: adaptive estimation

In this section we discuss some implications of our results relating to adaptive estimation. Let $\bar{\beta} > 1/2$ be the smoothness of $\theta \in \ell_2$. Then for any $1/2 < \beta_0 < \bar{\beta}$ the parameter θ belongs to a subspace Θ_{β_0}. Define the quantities

$$\mathrm{bias}_\theta(\beta) = \sum_{i=1}^{\infty} \frac{n^{-2}\theta_i^2}{(\tau_i^2(\beta)+n^{-1})^2} \quad \text{and} \quad \mathrm{var}_\theta(\beta) = \mathrm{E}_\theta \sum_{i=1}^{\infty} \frac{n^{-1}\tau_i^4(\beta)\xi_i^2}{(\tau_i^2(\beta)+n^{-1})^2},$$

which are in fact the bias and variance terms of the risk $\mathrm{E}_\theta\|\theta-\hat{\theta}(\beta)\|^2$ of the Bayes estimator $\hat{\theta}(\beta)$ corresponding to the prior $\pi(\beta)$ (see also Example 2).

If we substitute the parameter β in the Bayes estimator (3) by its estimate $\widehat{\beta}$, we get the estimator $\widehat{\theta}(\widehat{\beta})$ with components

$$\widehat{\theta}_i \equiv \widehat{\theta}_i(\widehat{\beta}) = \frac{\tau_i^2(\widehat{\beta})X_i}{\tau_i^2(\widehat{\beta})+n^{-1}}. \tag{15}$$

This estimator does not depend on the smoothness parameter of the Sobolev class. Define the risk of this estimator:

$$R(\widehat{\theta}(\widehat{\beta}),\theta) = \mathrm{E}_\theta\|\widehat{\theta}(\widehat{\beta})-\theta\|^2. \tag{16}$$

Then the following theorem holds true.

Theorem 4. *Let $\theta \in \Theta_{\beta_0}$, $\beta_0 < \bar{\beta}$ and Condition A be fulfilled for any $\delta_A \in (0,1/2]$. Then*

$$\lim_{n\to\infty} n^{2\beta_0/(2\beta_0+1)} R(\widehat{\theta}(\widehat{\beta}),\theta) \le C(\beta_0,\bar{\beta}),$$

where $C(\beta_0,\bar{\beta}) = A_\theta\Big([\bar{\beta}+\beta_0(2\bar{\beta}+1)/(2\beta_0+1)]/2\Big)(2\bar{\beta})^{2\bar{\beta}/(2\bar{\beta}+1)}$.

Proof. Notice that from the trivial inequality $(x+y)^2 \le 2x^2 + 2y^2$ we get

$$
\begin{aligned}
R(\widehat{\theta}, \theta) &= \mathrm{E}_\theta \|\widehat{\theta}(\widehat{\beta}) - \theta\|^2 \\
&= \mathrm{E}_\theta \sum_{i=1}^{\infty} \left(n^{-1/2} \xi_i \frac{\tau_i^2(\widehat{\beta})}{\tau_i^2(\widehat{\beta}) + n^{-1}} - \frac{n^{-1}\theta_i}{\tau_i^2(\widehat{\beta}) + n^{-1}} \right)^2 \\
&\le 2\mathrm{E}_\theta \sum_{i=1}^{\infty} n^{-1} \xi_i^2 \left(\frac{\tau_i^2(\widehat{\beta})}{\tau_i^2(\widehat{\beta}) + n^{-1}} \right)^2 + 2\mathrm{E}_\theta \sum_{i=1}^{\infty} \frac{n^{-2}\theta_i^2}{(\tau_i^2(\widehat{\beta}) + n^{-1})^2} \\
&= 2\mathrm{E}_\theta \sum_{i=1}^{\infty} \xi_i^2 \frac{n}{(n + i^{2\widehat{\beta}+1})^2} + 2\mathrm{E}_\theta \sum_{i=1}^{\infty} \theta_i^2 \frac{i^{4\widehat{\beta}+2}}{(n + i^{2\widehat{\beta}+1})^2} \equiv 2\mathrm{E}_\theta R_1 + 2\mathrm{E}_\theta R_2.
\end{aligned}
$$

In order to prove the theorem we have to show that for $k = 1, 2$

$$
\lim_{n\to\infty} n^{2\beta_0/(2\beta_0+1)} \mathrm{E}_\theta R_k = 0.
$$

Consider first the term R_1. Let $\beta' = [(\beta_0 + \bar{\beta})/2]_{S_n}$, $\delta_n = c/\log n$ as in Theorem 1 and $U_1 = (1/2, \beta' - \delta_n)$, $U_2 = [\beta' - \delta_n, +\infty)$. Then $\mathrm{E}_\theta R_1 = \mathrm{E}_\theta R_1 \mathbf{1}\{\widehat{\beta} \in U_1\} + \mathrm{E}_\theta R_2 \mathbf{1}\{\widehat{\beta} \in U_2\}$.

Using the Hölder inequality we get

$$
\mathrm{E}_\theta R_1 \mathbf{1}\{\widehat{\beta} \in U_k\} \le \left(\mathrm{E}_\theta \left(\sum_{i=1}^{\infty} \xi_i^2 \frac{n}{(n + i^{2\widehat{\beta}+1})^2} \right)^2 \right)^{1/2} \Big(\mathrm{P}_\theta\{\widehat{\beta} \in U_k\} \Big)^{1/2}, \quad k = 1, 2. \tag{17}
$$

Since $\beta' = [(\beta_0 + \bar{\beta})/2]_{S_n} < \bar{\beta}$ we have $A_\theta(\beta') < \infty$ and we can apply Theorem 1. There exists an integer $N = N(\beta_0, \bar{\beta}, \theta, C_2)$ such that for any $C > 0$ there exists a positive $C_0 = C_0(\beta_0, \bar{\beta}, C_2, C)$ such that for all $c > C_0$ we have

$$
\mathrm{P}_\theta\{\widehat{\beta} \in U_1\} \equiv \mathrm{P}_\theta\{\widehat{\beta} \le \beta' - \delta_n\} < C_1 \exp\{-C n^{1/(\beta_0 + \bar{\beta} + 1)}\} \quad \forall\, n > N. \tag{18}
$$

Notice now that for any $\widehat{\beta} > 1/2$ we have

$$
\begin{aligned}
\mathrm{E}_\theta \left(\sum_{i=1}^{\infty} \xi_i^2 \frac{n}{(n + i^{2\widehat{\beta}+1})^2} \right)^2 &< \mathrm{E}_\theta \left(\sum_{i=1}^{\infty} \xi_i^2 \frac{n}{(n + i^2)^2} \right)^2 \\
&= \mathrm{E}_\theta \left(\sum_{i=1}^{\infty} \xi_i^4 \frac{n^2}{(n+i^2)^4} + 2 \sum_{i=1}^{\infty} \sum_{j=i+1}^{\infty} \xi_i^2 \xi_j^2 \frac{n^2}{(n+i^2)^2 (n+j^2)^2} \right)^2 \\
&< 3 \sum_{i=1}^{\infty} \frac{n^2}{(n+i^2)^4} + 2n^2 \left(\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{1}{(n+i^2)^2} \right) \\
&= 3B(2,4,0) n^{-3/2} + 2B^2(2,2,0) n^{-1} + \phi_n,
\end{aligned}
$$

where the last inequality follows from Lemma 8, $|\phi_n| \le D n^{-2}$, D is some positive constant. Using this estimate, (18), and (17) we get that there exists an integer $N = N(\beta_0, \bar{\beta}, \theta, C_2)$ such that for all $n > N$

$$
\mathrm{E}_\theta R_1 \mathbf{1}\{\widehat{\beta} \in U_1\} \le 2\sqrt{2} B(2,2,0) \sqrt{C_1} n^{-1/2} \exp\{-\frac{1}{2} C n^{1/(\beta_0 + \bar{\beta} + 1)}\}. \tag{19}
$$

Let us estimate now the term $\mathrm{E}_\theta R_1 \mathbf{1}\{\widehat{\beta} \in U_2\}$. Since $\widehat{\beta} > \beta' - \delta_n$ we apply Lemma 8 and obtain

$$
\begin{aligned}
\mathrm{E}_\theta R_1 \mathbf{1}\{\widehat{\beta} \in U_2\} &\le \mathrm{E}_\theta \sum_{i=1}^{\infty} \xi_i^2 \frac{n}{(n + i^{2(\beta'-\delta_n)+1})^2} \\
&= \sum_{i=1}^{\infty} \frac{n}{(n + i^{2(\beta'-\delta_n)+1})^2} = B(2\beta' - 2\delta_n + 1, 2, 0) n^{-2(\beta'-\delta_n)/(2(\beta'-\delta_n)+1)} + \phi_n',
\end{aligned}
$$

where $|\phi_n'| \le D' n^{-1}$. It follows from $\beta' > \beta_0$ that $\beta' - \delta_n > \beta_0$ for sufficiently large n and $n^{-2(\beta'-\delta_n)/(2(\beta'-\delta_n)+1)} < n^{-2\beta_0/(2\beta_0+1)}$. Consequently,

$$
n^{2\beta_0/(2\beta_0+1)} \mathrm{E}_\theta R_1 \mathbf{1}\{\widehat{\beta} \in U_2\} \to 0, \quad n \to \infty.
$$

Combining this relation with (19) we get

$$
n^{2\beta_0/(2\beta_0+1)} \mathrm{E}_\theta R_1 \to 0, \quad n \to \infty. \tag{20}
$$

Consider now the term R_2. Let $\mathrm{E}_\theta R_2 = \mathrm{E}_\theta R_2 \mathbf{1}\{\widehat{\beta} \in U_3\} + \mathrm{E}_\theta R_2 \mathbf{1}\{\widehat{\beta} \in U_4\}$, where $U_3 = (1/2, \bar{\beta} + \delta_A]$, $U_4 = (\bar{\beta} + \delta_A, +\infty)$. Set $\beta'' = [(\bar{\beta} + \beta_0(2\bar{\beta}+1)/(2\beta_0+1))/2]_{S_n}$. Then $A_\theta(\beta'') < \infty$ and

$$
\begin{aligned}
\mathrm{E}_\theta R_2 \mathbf{1}\{\widehat{\beta} \in U_3\} &= \mathrm{E}_\theta \left\{ \sum_{i=1}^{\infty} \frac{i^{4\widehat{\beta}+2-2\beta''}}{(n + i^{2\widehat{\beta}+1})^2} i^{2\beta''} \theta_i^2 \mathbf{1}\{\widehat{\beta} \in U_2\} \right\} \\
&\le A_\theta(\beta'') \mathrm{E}_\theta \left\{ \max_i \frac{i^{4\widehat{\beta}+2-2\beta''}}{(n + i^{2\widehat{\beta}+1})^2} \mathbf{1}\{\widehat{\beta} \in U_2\} \right\} \\
&\le A_\theta(\beta'') \max_i \frac{i^{4(\bar{\beta}+\delta_A)+2-2\beta''}}{(n + i^{2(\bar{\beta}+\delta_A)+1})^2} \le A_\theta(\beta'') D n^{-2\beta''/(2\bar{\beta}+2\delta_A+1)} + \psi_n,
\end{aligned}
$$

where the last estimate follows from Lemma 8, $D = (2\bar{\beta})^{2\bar{\beta}/(2\bar{\beta}+1)}$ upper bounds the constant $D = D(2\bar{\beta} + 2\delta_A + 1, 2, 4(\bar{\beta} + \delta_A) + 2 - 2\beta'')$, $|\psi_n| \le D' n^{-(2\beta''+1)/(2\bar{\beta}+2\delta_A+1)}$, $D' > 0$ is some constant. Notice that for all n

$$
n^{-2\beta''/(2\bar{\beta}+2\delta_A+1)} = \left(n^{-2\beta_0/(2\beta_0+1)} \right)^{(2\bar{\beta}+1)/(2\bar{\beta}+2\delta_A+1)}.
$$

Tending δ_A to zero, we get

$$
n^{2\beta_0/(2\beta_0+1)} \mathrm{E}_\theta R_2 \mathbf{1}\{\widehat{\beta} \in U_3\} \to C(\beta_0, \bar{\beta}), \quad n \to \infty. \tag{21}
$$

Now, using the Hölder inequality and Theorem 2 it is easy to show that

$$
\begin{aligned}
\mathrm{E}_\theta R_2 \mathbf{1}\{\widehat{\beta} \in U_4\} &\le \left(\mathrm{E}_\theta \left(\sum_{i=1}^{\infty} \theta_i^2 \frac{i^{4\widehat{\beta}+2}}{(n + i^{2\widehat{\beta}+1})^2} \right)^2 \right)^{1/2} \left(\mathrm{P}_\theta\{\widehat{\beta} \in U_4\} \right)^{1/2} \\
&\le \|\theta\|^2 \left(\mathrm{P}_\theta\{\widehat{\beta} > \bar{\beta} + \delta_A\} \right)^{1/2} = \|\theta\|^2 |S_n| \exp\left\{ -\frac{1}{2}(C_A - K_2(\bar{\beta})) n^{1/(2\bar{\beta}+3)} \right\}
\end{aligned}
$$

for $n \geq N_A$. Thus this term decreases exponentially for properly chosen grid $|S_n|$. Combining this estimate with (21) and tending δ_A to zero we get

$$n^{2\beta_0/(2\beta_0+1)}\mathrm{E}_\theta R_2 \to C(\beta_0, \bar{\beta}), \quad n \to \infty,$$

and prove the theorem. $\square$

7 Appendix

In this section we gather several technical lemmas which are used in the proofs of the main results.

Let b_+ denote the nonnegative part of b. A version of the following auxiliary result is contained in Freedman (1999). As compared to Lemma 2 from Freedman (1999), our lemma below provides also bounds for the second order terms suitable for our purposes.

Lemma 8. *Suppose $0 < p < \infty$, $0 < q < \infty$, $0 \leq r < \infty$ and $pq > r + 1$. Let $\gamma_n \to \infty$ as $n \to \infty$. Then*

$$\sum_{i=1}^{\infty} \frac{i^r}{(\gamma_n + i^p)^q} = B(p,q,r)\gamma_n^{(1+r)/p-q} + \phi_n \,,$$

$$\max_{i\in\mathbb{N}} \frac{i^r}{(\gamma_n + i^p)^q} = D(p,q,r)\gamma_n^{(r/p)-q} + \psi_n \,,$$

where $\phi_n = \phi_n(p,q,r)$ and $\psi_n = \psi_n(p,q,r)$ are such that $|\phi_n| \leq D(p,q,r)\gamma_n^{-q+r/p}$, $|\psi_n| \leq C(p,q,r)\gamma_n^{-q+\frac{(r-1)_+}{p}}$ for some constant $C(p,q,r) > 0$, $B(p,q,r) = \int_0^\infty \frac{u^r du}{(1+u^p)^q}$ is defined by (4) and $D(p,q,r) = r^{r/p}(pq-r)^{q-(r/p)}(pq)^{-q} = \big(1 - \frac{r}{pq}\big)^q\big(\frac{pq}{r} - 1\big)^{-r/p}$, with the convention $0^0 = 1$.

Remark 13. Notice that if $r \geq 1$ then $0 \leq D(p,q,r) \leq 1$.

Proof. Denote $g(u) = \frac{u^r}{(\gamma_n+u^p)^q}$, $u > 0$. Function $g(u)$ is increasing on $u \in [0, u_{max}]$ en decreasing on $[u_{max}, \infty)$ with $u_{max} = \big(r\gamma_n/(pq-r)\big)^{1/p}$. Therefore,

$$\int_0^\infty \frac{u^r}{(1+u^p)^q} - g(u_{max}) \leq \sum_{i=1}^{\infty} \frac{i^r}{(\gamma_n + i^p)^q} \leq \int_0^\infty \frac{u^r}{(1+u^p)^q} + g(u_{max}) \,,$$

with $g(u_{max}) = D(p,q,r)\gamma_n^{(r/p)-q}$, which establishes the first relation. To prove the second relation, we first compute $g'(u) = \frac{r\gamma_n u^{r-1} - (pq-r)u^{p+r-1}}{(\gamma_n+u^p)^{q+1}}$ and then evaluate

$$\max_{u\geq 1}|g'(u)| \leq \max\Big\{\max_{u\geq 1}\Big\{\frac{r\gamma_n u^{r-1}}{(\gamma_n+u^p)^{q+1}}\Big\}, \max_{u\geq 1}\Big\{\frac{(pq-r)u^{p+r-1}}{(\gamma_n+u^p)^{q+1}}\Big\}\Big\} \leq C(p,q,r)\gamma_n^{-q+\frac{(r-1)_+}{p}}$$

for some constant $C(p,q,r) > 0$. Finally, using this bound and the unimodality of the function $g(u)$ on $[0,\infty)$, we obtain

$$\Big|g(u_{max}) - \max_{i\in\mathbb{N}} \frac{i^r}{(\gamma_n + i^p)^q}\Big| \leq \max_{u\geq 1}|g'(u)| \leq C(p,q,r)\gamma_n^{-q+\frac{(r-1)_+}{p}} \,,$$

which completes the proof of the lemma. □

The following short lemma that we give without proof follows directly from the properties of Beta and Gamma functions.

Lemma 9. *Let $p, q, r > 0$, $pq > r + 1$. Then for $q = 1, 2$ we have*

$$B(p,1,r) = \frac{\pi}{p \sin\Big(\pi(r+1)/p\Big)}, \quad B(p,2,r) = \frac{\pi(p-r-1)}{p^2 \sin\Big(\pi(r+1)/p\Big)}.$$

Moreover, if $\beta \geq \bar\beta > 1/2$, we have the following bounds for the function B:

$$(2\bar\beta)^{-1} \leq B(2\beta+1, 1, 2(\beta-\bar\beta)) \leq \pi\Big((2\bar\beta+1)\sin\big(\pi/(2\bar\beta+1)\big)\Big)^{-1},$$
$$(2\beta+1)^{-1} \leq B(2\beta+1, 2, 2(\beta-\bar\beta)) \leq \pi\Big((2\bar\beta+1)\sin\big(\pi/(2\bar\beta+1)\big)\Big)^{-1}.$$

Lemma 10. *Let $a_i(\beta, \beta_0)$ be defined by (6) and $A_n^2 = \sum_{i=1}^\infty a_i^2(\beta, \beta_0)$. There exists an integer $N = N(\beta, \beta_0)$ such that for any $n > N$*

$$\frac{1}{16} n^{1/(2\gamma+1)} \leq n^{-2} A_n^2 < B(2\gamma+1, 2, 0) n^{1/(2\gamma+1)}, \tag{22}$$

where $\gamma = \min(\beta, \beta_0)$ and B is defined by (4).

Proof. Recalling (6), we have

$$n^{-2} \sum_{i=1}^\infty a_i^2(\beta, \beta_0) = n^{-2} \sum_{i=1}^\infty \frac{\tau_i^4(\beta_0) + \tau_i^4(\beta) - 2\tau_i^2(\beta_0)\tau_i^2(\beta)}{(\tau_i^2(\beta) + n^{-1})^2 (\tau_i^2(\beta_0) + n^{-1})^2} \equiv S_1 + S_2 - 2S_3,$$

where

$$S_1 = \sum_{i=1}^\infty \frac{n^2 i^{4\beta+2}}{(n + i^{2\beta+1})^2 (n + i^{2\beta_0+1})^2}, \quad S_2 = \sum_{i=1}^\infty \frac{n^2 i^{4\beta_0+2}}{(n + i^{2\beta+1})^2 (n + i^{2\beta_0+1})^2},$$
$$S_3 = \sum_{i=1}^\infty \frac{n^2 i^{2(\beta+\beta_0+1)}}{(n + i^{2\beta+1})^2 (n + i^{2\beta_0+1})^2}.$$

By Lemma 8, we evaluate for any β, β_0

$$S_1 < \sum_{i=1}^\infty \frac{n^2}{(n + i^{2\beta_0+1})^2} = B(2\beta_0+1, 2, 0) n^{1/(2\beta_0+1)} + \phi_{n,1}, \tag{23}$$

$$S_2 < \sum_{i=1}^\infty \frac{n^2}{(n + i^{2\beta+1})^2} = B(2\beta+1, 2, 0) n^{1/(2\beta+1)} + \phi_{n,2}, \tag{24}$$

where $\phi_{n,1} = \phi_n(2\beta_0+1, 2, 0)$, $\phi_{n,2} = \phi_n(2\beta+1, 2, 0)$. Thus we have the upper bound for (22).

Denote $N_\beta = \lfloor n^{1/(2\beta+1)} \rfloor$, $N_{\beta_0} = \lfloor n^{1/(2\beta_0+1)} \rfloor$, $N_{\beta,\beta_0} = \lfloor n^{1/(\beta+\beta_0+1)} \rfloor$.

If $\beta > \beta_0$ then

$$S_1 = \sum_{i=1}^{\infty} \frac{n^2 i^{4\beta+2}}{(n+i^{2\beta+1})^2(n+i^{2\beta_0+1})^2} \geq \sum_{i=N_\beta+1}^{N_{\beta_0}} \frac{n^2 i^{4\beta+2}}{(2n)^2(2i^{2\beta+1})^2} \geq \frac{1}{16}(n^{1/(2\beta_0+1)} - n^{1/(2\beta+1)}).$$

Taking into account (24) we can neglect S_2 because it is smaller in order than $n^{1/(2\beta_0+1)}$. Let us estimate S_3. We have

$$S_3 = \sum_{i=1}^{\infty} n^2 \frac{i^{2(\beta+\beta_0+1)}}{(i^{2\beta_0+1}+n)^2(i^{2\beta+1}+n)^2} \leq \sum_{i=1}^{N_{\beta,\beta_0}} \frac{n^2 i^{2(\beta+\beta_0+1)}}{n^2 \cdot n^2} + \sum_{i=N_{\beta,\beta_0}+1}^{\infty} \frac{n^2 i^{2(\beta+\beta_0+1)}}{i^{4(\beta+\beta_0+1)}}$$
$$= C(\beta,\beta_0) n^{1/(\beta+\beta_0+1)}(1+o(1)),$$

where $C(\beta,\beta_0)$ is some constant. Combining the estimates for S_1, S_2, and S_3 we get the lower bound of (22) for the case $\beta > \beta_0$. The proof of (22) for $\beta < \beta_0$ is analogous. □

At the end of this section an alternative proof of Lemma 7 is given. It is based on the following result which is essentially contained in Cavalier, Golubev, Picard and Tsybakov (2002) (see Lemma 2 in that paper, cf. also Lemma 4 in Freedman (1999)). We provide the proof for the completeness.

Lemma 11. *Let ξ_i's be independent standard normal variables, $a = \{a_i\}_{i\in\mathbb{N}} \in \ell_2$, i.e. $\|a\|^2 = \sum_{i=1}^{\infty} a_i^2 < \infty$ and $c_a = \sup_{i\in\mathbb{N}} |a_i|/\|a\|$. Then for any $u > 0$*

$$\mathrm{P}\Big(\Big|\sum_{i=1}^{\infty} a_i(\xi_i^2-1)\Big| > u\Big) \leq 2\exp\Big\{-\frac{u^2}{4\|a\|^2\big(1+uc_a\|a\|^{-1}\big)}\Big\}.$$

Proof. Denote $V = \sum_{i=1}^{\infty} a_i(\xi_i^2-1)$. Recall that for $\xi \sim \mathcal{N}(0,1)$ and $\lambda < 1/2$,

$$\mathrm{E}\exp\{\lambda(\xi^2-1)\} = (1-2\lambda)^{-1/2}e^{-\lambda}.$$

Using this relation, the Markov inequality and the expansion $\log(1-x) = -\sum_{k=1}^{\infty}\frac{x^k}{k}$, we have that

$$\mathrm{P}(V > u) \leq e^{-\lambda u}\exp\Big\{-\sum_{i=1}^{\infty}\Big(\lambda a_i + \frac{1}{2}\log(1-2\lambda a_i)\Big)\Big\}$$

for λ such that $\lambda a_i < 1/2$ for all $i = 1, 2, \ldots$, which is in turn true for all $\lambda < (2c_a\|a\|)^{-1}$.

Now we evaluate

$$
\begin{aligned}
-\sum_{i=1}^{\infty}\Big(\lambda a_i+\frac{1}{2}\log(1-2\lambda a_i)\Big) &= \sum_{i=1}^{\infty}\Big(-\lambda a_i+\sum_{k=1}^{\infty}\frac{(2\lambda a_i)^k}{2k}\Big)\\
&= \sum_{i=1}^{\infty}\sum_{k=2}^{\infty}\frac{(2\lambda a_i)^k}{2k}\le\sum_{k=2}^{\infty}\frac{(2\lambda)^k\|a\|^k}{2k}\sum_{i=1}^{\infty}\frac{|a_i|^k}{\|a\|^k}\\
&\le \sum_{k=2}^{\infty}\frac{(2\lambda)^k\|a\|^k}{2k}c_a^{k-2}=\frac{1}{2c_a^2}\sum_{k=2}^{\infty}\frac{(2\lambda c_a\|a\|)^k}{k}\\
&= -\frac{\log\big(1-2\lambda c_a\|a\|\big)}{2c_a^2}-\frac{\lambda\|a\|}{c_a},
\end{aligned}
$$

which leads to

$$
\mathrm{P}(V>u)\le\exp\Big\{-\frac{\log\big(1-2\lambda c_a\|a\|\big)}{2c_a^2}-\frac{\lambda\|a\|}{c_a}-\lambda u\Big\}=\exp\big\{g_u(\lambda)\big\}
$$

for all $\lambda<(2c_a\|a\|)^{-1}$. Minimizing the function $g_u(\lambda)$ defined above, we derive the optimal

$$
\lambda_o=\lambda_o(u)=\frac{u}{2\|a\|(\|a\|+uc_a)}\,.
$$

For each $u>0$, we obviously have that $\lambda_o<(2c_a\|a\|)^{-1}$ so we can use this λ in the above bound:

$$
\mathrm{P}(V>u)\le\exp\big\{g_u(\lambda_o)\big\}=\exp\Big\{\frac{1}{2c_a^2}\Big(\log\big(1+uc_a\|a\|^{-1}\big)-uc_a\|a\|^{-1}\Big)\Big\}\,.
$$

Finally using the elementary inequality for any $x\ge 0$

$$
\log(1+x)-x\le-\frac{x^2}{2(1+x)}\,,
$$

we obtain

$$
\mathrm{P}(V>u)\le\exp\Big\{\frac{1}{2c_a^2}\frac{u^2c_a^2\|a\|^{-2}}{2(1+uc_a\|a\|^{-1})}\Big\}=\exp\Big\{-\frac{u^2}{4\|a\|^2\big(1+uc_a\|a\|^{-1}\big)}\Big\}\,.
$$

The lemma is proved. □

Lemma 12. *1. Let Condition A be fulfilled,* $\bar\beta+\delta<\beta\le\bar\beta+1$ *and the constants* δ_A, K_A, N_A *be defined in Condition A. Then there exist an integer* $N=N(\theta,\delta_A,K_A,N_A)$ *such that*

$$
\mathrm{P}_\theta\{\widehat\beta(S_n)=\beta\}\le 2\exp\Big\{-\frac{C}{8\sqrt{B}}n^{(2\bar\beta-1)/(2(2\bar\beta+3)^2)}\Big\}+\frac{n^{-1/(4\bar\beta+6)}}{\mu\sqrt{C_A}}\exp\Big\{-\frac{\mu^2C_A}{2}n^{1/(2\bar\beta+3)}\Big\}
$$

for all $n\ge N$, $0<\delta<1-2^{-1/(2\bar\beta+5/2)}$, $0<\mu<1/2-4K_2(\bar\beta)/(2C_A)$, *where* $C_A=\min\{K_A,\delta K/8\}$, $C=C(\bar\beta,K_A,\delta,\mu)=(1-2\mu)C_A-K_2(\bar\beta)>0$, $K_2(\bar\beta)=2+(2\bar\beta)^{-1}$, *and* $B=\pi\big((2\bar\beta+1)\sin(\pi/(2\bar\beta+1))\big)^{-1}$.

2. If $\beta > \bar\beta + 1$ then there exist an integer $N = N(\theta, K)$ such that for all $n \ge N$

$$\mathrm{P}_\theta\{\widehat\beta(S_n) = \beta\} \le 2\exp\Big\{-\frac{C}{8\sqrt{B}} n^{1/(2(2\bar\beta+1))}\Big\} + \frac{n^{-1/(4\bar\beta+2)}}{\mu\sqrt{C_A}} \exp\Big\{-\frac{\mu^2 C_A}{2} n^{1/(2\bar\beta+1)}\Big\},$$

where $0 < \delta < 1 - 4^{-1/(2\bar\beta+1)}$, $0 < \mu < 1/2 - 4K_2(\bar\beta)/(2C_A)$, where $C_A = \delta K/8$, $C = C(\bar\beta, \mu, K) = (1-2\mu)C_A - K_2(\bar\beta) > 0$, $K_2(\bar\beta) = 2+(2\bar\beta)^{-1}$, and $B = \pi\big((2\bar\beta+1)\sin(\pi/(2\bar\beta+1))\big)^{-1}$.

Proof. For the sake of brevity, in this proof we use denote $a_i = a_i(\beta, \beta')$, $b_i = b_i(\beta, \beta')$, $\beta' = [\beta_A(2\beta+1)/(2\beta_A+1)]_{S_n}$. According to the formulas (6) and (7), we have that $b_i^{-1} = 1 + a_i(\tau_i^2(\beta_A) + n^{-1})$. Therefore, the elementary inequality $\log(1+x) \le x$, $x > -1$, implies that

$$\sum_{i=1}^\infty \big(\log b_i + n^{-1} a_i\big) = -\sum_{i=1}^\infty \big(\log b_i^{-1} - n^{-1} a_i\big) \ge -\sum_{i=1}^\infty a_i \tau_i^2(\beta_0)\,.$$

Note that $a_i(\beta, \beta') > 0$ for $\beta' < \beta$. Using (8) and the last relation, we get for any $\mu > 0$

$$\begin{aligned}
\mathrm{P}_\theta\{\widehat\beta(S_n) = \beta\} &= \mathrm{P}_\theta\{Z_n(\beta, \beta'') \le 0 \,\forall\, \beta'' \in S_n\} \le \mathrm{P}_\theta\{Z_n(\beta, \beta') \le 0\} \\
&= \mathrm{P}_\theta\Big\{-\sum_{i=1}^\infty a_i(\beta,\beta') X_i^2 \ge \sum_{i=1}^\infty \log b_i(\beta,\beta')\Big\} \\
&= \mathrm{P}_\theta\Big\{-\sum_{i=1}^\infty a_i(\beta,\beta')\Big(\theta_i^2 + \frac{2\theta_i\xi_i}{\sqrt{n}} + \frac{\xi_i^2}{n}\Big) \ge \sum_{i=1}^\infty \log b_i(\beta,\beta')\Big\} \\
&= \mathrm{P}_\theta\Big\{-\sum_{i=1}^\infty a_i\Big(\frac{\xi_i^2 - 1}{n} + \frac{2\theta_i\xi_i}{\sqrt{n}}\Big) \ge \sum_{i=1}^\infty a_i(\theta_i^2 - \tau_i^2(\beta'))\Big\} \\
&\le \mathrm{P}_\theta\Big\{-\sum_{i=1}^\infty \frac{a_i(\xi_i^2-1)}{n} \ge \sum_{i=1}^\infty a_i\Big(\frac{\theta_i^2}{2} - \tau_i^2(\beta')\Big) - \mu\sum_{i=1}^\infty a_i\theta_i^2\Big\} \\
&\qquad + \mathrm{P}_\theta\Big\{\sum_{i=1}^\infty \frac{2a_i\theta_i\xi_i}{\sqrt{n}} \le -\mu\sum_{i=1}^\infty a_i\theta_i^2\Big\}\,.
\end{aligned} \tag{25}$$

Let $\bar\beta + \delta_A < \beta \le \bar\beta + 1$. Using Lemma 5 and Lemma 6, we get that there exists an integer $N_1 = N_1(\theta, \delta_A, K_A, N_A)$ such that for all $n > N_1$

$$\begin{aligned}
n^{-1}u \equiv \sum_{i=1}^\infty a_i\Big(\frac{\theta_i^2}{2} - \tau_i^2(\beta')\Big) - \mu\sum_{i=1}^\infty a_i\theta_i^2 &= (1-2\mu)\sum_{i=1}^\infty a_i\Big(\frac{\theta_i^2}{2} - \tau_i^2(\beta')\Big) - 2\mu\sum_{i=1}^\infty a_i\tau_i^2(\beta') \\
&\ge C n^{\frac{1}{2\bar\beta+3}},
\end{aligned}$$

where $C = C(\bar\beta, K_A, \delta, \mu) = (1-2\mu)C_A - K_2(\bar\beta)$ is a positive constant for sufficiently small μ with $C_A = \min\{K_A, \delta K/8\}$, $0 < \delta < 1 - 2^{-1/(2\bar\beta+5/2)}$ (see also Remark 9). Using Lemma 10, we can estimate $\|a\|^2 = \sum_{i=1}^\infty a_i^2 \le B(2\beta'+1, 2, 0)n^{2+\frac{1}{2\beta'+1}}$ and then upper bound the constant $B(2\beta'+1,2,0)$ by $B=\pi\big((2\bar\beta+1)\sin(\pi/(2\bar\beta+1))\big)^{-1}$ using Lemma 9. Taking into account that $c_a\le 1$ and $\|a\|$ is smaller in order than u, we apply Lemma 11 to bound the first term in the right hand side of inequality (25):

$$\mathrm{P}_\theta\Big\{-\sum_{i=1}^\infty \frac{a_i(\xi_i^2-1)}{n}\ge un^{-1}\Big\}\le 2\exp\Big\{-\frac{u^2}{4\|a\|^2(1+uc_a\|a\|^{-1})}\Big\}$$
$$\le 2\exp\Big\{-\frac{u}{8\|a\|}\Big\}\le 2\exp\Big\{-\frac{C}{8\sqrt{B}}n^{(2\bar\beta-1)/(2(2\bar\beta+3)^2)}\Big\}$$

for all $n>N_2=N_2(\theta,\delta_A,K_A,N_A)$.

As to the second term in the right hand side of (25), notice first it is of the form $\mathrm{P}\{\eta\ge x\}$, with $\eta\sim\mathcal{N}(0,\sigma^2)$, $\sigma^2=n^{-1}\sum_{i=1}^\infty a_i^2\theta_i^2$, $x=\frac{\mu}{2}\sum_{i=1}^\infty a_i\theta_i^2$. Since $|a_i|\le n$ for all $i\in\mathbb{N}$, we have $\sum_{i=1}^\infty a_i\theta_i^2\ge n^{-1}\sum_{i=1}^\infty a_i^2\theta_i^2$, and consequently

$$\frac{x^2}{\sigma^2}=\frac{\mu^2\Big(\sum_{i=1}^\infty a_i\theta_i^2\Big)^2}{4n^{-1}\sum_{i=1}^\infty a_i^2\theta_i^2}\ge\frac{\mu^2}{4}\sum_{i=1}^\infty a_i\theta_i^2\,.$$

From Lemma 6 it follows that there exists an integer $N_3=N_3(\theta,\delta_A,K_A,N_A)$ such that for all $n\ge N_3$

$$\sum_{i=1}^\infty a_i\theta_i^2\ge 2C_An^{1/(2\bar\beta+3)}.$$

Thus we obtain $\frac{x^2}{\sigma^2}\ge\mu^2C_An^{1/(2\bar\beta+3)}$ for all $n\ge N_3$. Applying the elementary inequality $\mathrm{P}\{\eta\ge x\}\le\frac{\sigma}{x}\exp\big\{-\frac{x^2}{2\sigma^2}\big\}$ and taking $N=\max(N_1.N_2.N_3)$ completes the proof of the first part. Similarly, we can prove the second part for $\beta>\bar\beta+1$ using the second part of Lemma 6. □

References

[1] Belitser, E. and Levit, B. (2003). On the empirical Bayes approach to adaptive filtering. *Math. Meth. Statist.* **12** 131–154.

[2] Belitser, E. and Ghosal, S. (2003). Adaptive Bayesian inference on the mean of an infinite dimensional normal distribution. *Ann. Statist.* **31** 536–559.

[3] Birgé, L. and Massart, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* **3** 203–268.

[4] Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384–2398.

[5] Cavalier, L. and Tsybakov, A.B. (2001). Penalized blockwise Stein's method, monotone oracles and sharp adaptive estimation. *Mathematical Methods of Statistics* **10** 247–282.

[6] Cavalier, L., Golubev, G.K., Picard, D. and Tsybakov, A.B. (2002) Oracle inequalities for inverse problems. *Ann. Statist.* **30** 843–874.

[7] Ibragimov, I. A. and Khasminski, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer, New York.

[8] Johnstone, I. (1999). *Function Estimation in Gaussian Noise: Sequence models.* Monograph draft, http://www-stat.stanford.edu/∼imj/

[9] Freedman, D. (1999). On the Bernstein-von Mises theorem with infinite dimensional parameters. *Ann. Statist.* **27** 1119–1140.

[10] Grama, I. and Nussbaum, M. (1998). Asymptotic equivalence for nonparametric generalized linear models. *Probab. Theory Related Fields* **111** 167–214.

[11] Ingster, Yu.I. adn Suslina I.A. (2003) *Nonparametric Goodness-Of-Fit Testing under Gaussian Models.* Springer, New York.

[12] Lepski, O. and Hoffmann, M. (2002). Random rates in anisotropic regression. *Ann. Statist.* **30** 325–396.

[13] Low, M.G. (1997). On nonparametric confidence intervals. *Ann. Statist.* **25** 2547–2554.

[14] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24** 2399–2430.

[15] Pinsker, M. S. (1980). Optimal filtration of square-integrable signals in Gaussian noise. *Problems of Information Transmission* **16** 120–133.

[16] Robins, J. and van der Vaart, A. (2004). Adaptive Nonparametric Confidence Sets. *Preprint.* Vrije Universiteit Amsterdam.

[17] Robbins, H. (1955). An empirical Bayes approach to statistics. *In: Proc. 3rd Berkeley Symp. on Math. Statist. and Prob.* **1** Berkeley, Univ. of California Press, 157–164.

[18] Tsybakov, A.B. (2004). *Introduction à l'estimation non-paramétrique.* Springer, Berlin.

Department of Mathematics
Utrecht University
P.O. Box 80010
3508 TA Utrecht
The Netherlands
E-mail: belitser@math.uu.nl

EURANDOM
P.O. Box 513
5600 MB Eindhoven
The Netherlands
E-mail: enikeeva@eurandom.tue.nl